TAMPERE UNIVERSITY OF TECHNOLOGY

Wenjie Bao

PROPAGATION-DEPENDENT FAIR RATE ALLOCATION IN HETEROGENEOUS CELLULAR SYSTEM

Master of Science Thesis

Examiner:Dr. Dmitri Moltchanov
Prof. Yevgeni Koucheryavy
Examiner and topic approved by the Faculty Council of the Faculty of Computing and Electrical Engineering on 8th October 2014

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master‘s Degree Programme in Infromation Technology
Bao, Wenjie: PROPAGATION-DEPENDENT FAIR RATE ALLOCATION IN HETEROGENEOUS CELLULAR SYSTEM
Master of Science Thesis, 50 pages
January 2015
Major: Communication Systems and Networks
Examiner: Dr. Dmitri Moltchanov
Keywords: HetNets, Cellular, Fair networks, MMF, spectral efficiency, priority

In this day and age, the cellular communication network is advancing at an amazing speed. Heterogeneous wireless access technologies would play an increasingly critical role in next generation wireless communication systems.

The ultimate objective of this thesis is to implement fair and priority-based rate allocation in heterogeneous cellular system using Max-Min fairness criterion. Two coefficients are introduced successfully in the proposed algorithm. Consequently, it accomplish the goal of resource allocation in terms of spectral efficiency and prioritization in heterogeneous cellular networks.

However, the implementation is simplified to prove the correctness of the algorithm. More realistic scenario should be taken into consideration. Besides, interference would affect the final optimal solution.

PREFACE

The research leading to this thesis has been carried out at the Department of Electronics Communications Engineering, Tampere University of Technology, Finland. It starts from June to October,2014 for completing this thesis under the guidance of my supervisors, Prof.Yevgeni Koucheryavy and Dr.Dmitri Moltchanov.

First of all, I would like to express my gratitude to my supervisors Prof. Yevgeni Koucheryavy and Dr.Dmitri Moltchanov for giving me the chance to study on this topic and their support during my thesis study even in their busy schedule. And, Dr.Dmitri Moltchanov always encourages and guides me when I became depressed after meeting problems. He kept meeting me once a week to make sure the schedule of my thesis and to answer my questions.

Even if I had many ups and downs, I believe I have learned a lot from this topic and my supervisors. I really appreciate the opportunity of Master degree study which is offered by Tampere University of Technology.

Finally, I would like to thank my parents and friends for supporting me during the study in Finland.

Wenjie Bao
Tampere, January 2015.

CONTENTS

LIST OF ABBREVIATIONS

QoS	Quality of Service
HetNet	Heterogeneous Networks
LTE	Lone-Term Evolution
GSM	Global System for Mobile Communications
CDMA	Code Division Multiple Access
WiMAX	Worldwide Interoperability for Microwave Access
OFDMA	Orthogonal Frequency Division Multiple Access
IP	Internet Protocol
RAT	Radio Access Technology
HSPA	High Speed Packet Access
NFP	Network Flow Problem
BS	Base Station
AMPL	AMPL modeling language
MMF	Max-Min Fairness
PF	Proportional Fairness

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

In this day and age, the cellular communication network is advancing at an amazing speed. Heterogeneous wireless access technologies would play an increasingly critical role in next generation wireless communication systems. The cellular system present various advantages in terms of service capacity, mobility support, and quality of service(QoS). A heterogeneous network(HetNet) is typically composed of multiple radio access technologies, architectures, transmission solutions, and base stations of varying transmission power[1]. Since the world enters the era of 'Big Data', the heterogeneous cellular system also need to evolve to meet the overwhelming demands for capacity and bandwidth. In this thesis, we propose an optimal solution to improve the bandwidth utilization.

1.1 Current state of cellular networks

A cellular network allows cellular subscribers to wander anywhere in the country and remains connected to the Public Switched Telephone Network (PSTN) via their mobile phones[2]. A cellular network has a hierarchical structure and it is formed by connecting the mobile phones, base station and mobile switching center. It is extensively acknowledged that cellular networks has experienced explosive growth in nearly several-decades. Nowadays, millions of people are allowed to connect to a cellular network even if they are moving.

The cellular network has gone through four main generations, which are 1G,2G,3G and 4G LTE. At the beginning, the phone was analogue and offered basic voice service with very low levels of spectral efficiency and security. After that, 2G started to use digital technology (GSM,CDMA) and had much better spectral efficiency, security and new features (text messages, low data rate service). In the third generation (3G), higher speed data service was provided. Also, first mobile broadband was created. 4G LTE is the current generation of cellular technology. LTE and WiMAX are the major two channels, however, LTE-Advanced and WiMAX2 can fully meet the requirement of 4G. Long-Term Evolution (LTE) is a standard for wireless communication of high-speed data transfer. LTE effectively increases the capacity and speed since it uses a different radio interface with core network improvements. On the other side, Worldwide Interoperability for Microwave Access (WiMAX) uses a new physical layer radio access technology called Orthogonal Frequency Division

Multiple Access (OFDM) for uplink and downlink[3]. Obviously, LTE-Advanced andWiMAX2 are major enhancements of the LTE and WiMAX.

More precisely, the aim of 4G LTE is to provide high speed transmission with data rate up to 20Mbps while simultaneously accommodating QoS features. It is all-IP packet-switched networks supporting mobile ultra-broadband access. The high speed of 4G LTE meets high performance streaming of multimedia content. Also, it makes the video conference possible. Personal Area Networks (PANs), body LANs, low power sensors networked applications and self-configuring ad hoc networks will be encompassed in the 'sphere' of 4G[4].

4G networks have several advantages and disadvantages. For one thing, 4G network has an amazing speed compared with the previous generation. People could experience a superior and uninterrupted connectivity. Also, 4G networks provide larger coverage than other systems, such as Wi-Fi which makes users to depend on hotspots in their nearby area. While, 4G could offer a coverage of 30 miles and more, as also overlapping network, people are guaranteed with continued connection. Another advantage of 4G networks is security for mobile devices. 4G networks provide sufficient privacy, security and safety. This advantage is especially beneficial for online business. On the other side, we could still find some shortcomings about 4G networks. Even he coverage of 4G networks is increasing rapidly, it is still not available in many countries and regions. And, 4G mobile networks require more antennas and transmitters, it could lead to more power consumption of battery.

At present, mobile internet is widely used to do business in all industries all over the world. It attracts people with flexible ways supported by mobile access network and devices. The mobile access to the internet, cloud-based services is growing rapidly. Hence, we need to make breakthroughs in the transformation of network infrastructure.

Between 2020 and 2030, next generation networks (5G) will be deployed. It will support much higher capacity with extremely low latency and response time. 5G radio access will be built upon both new radio access technologies (RAT) and evolved existing wireless technologies (LTE, HSPA, GSM and Wi-Fi) [5]. The current generation of mobile networks still transform the way people exchange information. In the future, we will focus on true human-centric and connected machine-centric networks.

1.2 Definition of heterogeneous networks and motivations

Currently, mobile internet has experienced a dramatically growth in demand of data capacity. In this situation, heterogeneous networks becomes an effective methods of expanding network capacity. Recall the definition above, A heterogeneous network(HetNet) is typically composed of multiple radio access technologies, architectures, transmission solutions, and base stations of varying transmission power[1]. In other words, HetNet is a hierarchical deployment that combines large (macro) cells with small cells using different radio access technologies in order to satisfy the exponential growth of wireless data. A simple illustration of HetNet is shown in Figure 1.1.

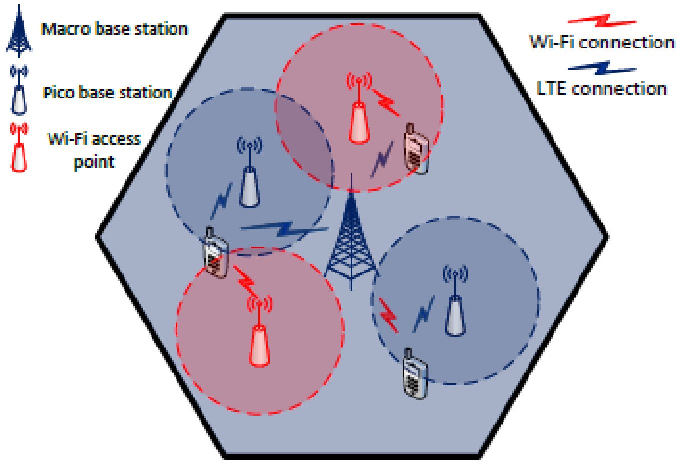

Figure 1.1: Simplified deployment example of a heterogeneous network [6].

With the people download more videos, transfer more data and access to the internet with their mobile devices, the data over the network is increasing astonishingly. Operators need to provide better data service, and make profits at the same time. Deploying a heterogeneous network is a approach to satisfy operators and customers' expectation of high data throughput with very low latency. To achieve this target, operators need to prepare networks for surging traffic demand. Besides, they should improve and densify their existing mobile broadband networks and use more integrated small cells in an optimal way.

Current wireless cellular networks are typically deployed as homogeneous networks using a macro-centric planned process[7]. A homogeneous cellular system is network in which the base stations are situated in specific positions. Also, the base stations all have similar transmission power levels, antenna patterns, receiver noise floors, and similar backhaul connectivity to the network. Furthermore, all base stations do not restrict access to user terminals in the network, and almost have the same number of user terminals in their coverage. All of the user terminals carry similar data flows with similar QoS requirements.

Nevertheless, radio link performance is approaching theoretical limits with LTE, *so the next performance breakthrough in wireless networks will come from the network topology.* Moreover, LTE-Advanced is mainly about increasing spectral efficiency per unit coverage. Even though air interface improvement could further maximize the benefits of advanced wireless communication research and fully utilize advanced signal processing techniques for higher spectral efficiency, we still need even larger capacity than what the air interface spectral efficiency improvement could provide. Hence, we need to come from a combination of technology solution, including the improvement at the radio link level, in particular. Heterogeneous networks are a fundamental technology behind those solutions.

Resource management and optimization are key means which will enable the efficient and effective operation of wireless networks, in order to minimize the impact of interference and maximize the overall network performance. The RATs in heterogeneous networks offer an additional method to achieve an efficient exploitation of the insufficient available radio resources. The selection of RATs becomes an important issue to achieve the expected result from the joint consideration of the heterogeneous characteristics offered by the available radio access networks. We should take various criteria into consideration when we choose the suitable RAT. Those criteria (such as ,service type, load conditions, cost, etc.) could fulfill the target of improving overall capacity, resource utilization and service quality.

As a consequence, we could use those criteria and technologies to get visibility into what is consuming bandwidth and causing problems. Bandwidth allocations needs to be controlled to ensure key applications get required bandwidth. It is also important to accelerate cloud, the Internet and video contents to reduce bandwidth consumption and improve end-user performance.

1.3 Research challenges in heterogeneous networks

There are several unsolved problems existing in the HetNet, and they could have a detrimental influence on adoption, utility and performance of heterogeneous networks. At the beginning, heterogeneous networks are less efficient and hard to deploy than homogenous networks at this moment since the lack of source integration, and because users or applications need to work around problems associated with the boundaries and the differences between the modalities. Network and resource management is normally organized by each modality, together with minimal support offered for enterprise-level cross-modality resource management. This leads to misallocation of resources and long response time to changing load or component failures.

Then, HetNet is more rigid, and provide a lower level of user or application control than homogenous networks. The experience of the network to users is inflexible since the poor resource management. Similarly, the lack of seamless connectivity could result in static pre-configuration of capacity allocations, prioritization, routes and so on. That is complex and difficult to change rapidly. The low level of integration between the separate modalities evidently constrains the choices available to application designers for QoS and bandwidth guarantees. Further, heterogeneous networks are fragile and fail more often than homogenous networks, due to the same poor resource management and static pre-configuration issues just described[8].

Another challenge is to achieve high performance with no requirement about being customized for the networks and technologies in use. In other words, users and applications are forced to cope with the heterogeneity of the network. The poor resource management and low flexibility of heterogeneous networks could result in consuming more power than before. It is complex to preserve the battery power of resource-constrained components because of the distributed network control and inefficient design. It seems impossible to optimize the use of end-to-end network capacity for power efficiency.

Lastly, HetNet is still less secure than homogenous networks. The increased system complexity makes it more difficult to distinguish the normal operation from attacks if there are multiple heterogeneous services or modalities. Heterogeneous networks should be more secure than homogenous networks, however, without appropriate architectural design, it is only a secure but the least robust subnetwork.

1.4 Problem statement

Based on the discussion above, the significance of heterogeneous cellular networks has attracted considerable attention. Nevertheless, the existing optimal solution has not fully met the requirement of fairness allocation of bandwidth in the previous study. Besides, fair rate allocation has not been sufficiently studied by far, and majority of the study only focuses on maximizing the upper bound throughput of heterogeneous networks.

In this thesis, we proposed an algorithm to fulfill fair rate allocation based on Max-min fairness criterion. To accomplish that, we regard the problem as a network flow problem first. More specifically, we establish two-tier heterogeneous networks with 30 subscribers uniformly random distributed. Each of these subscribers requires traffic demands from macro-LTE base stations and/or Wi-Fi base stations. Both macro-LTE and Wi-Fi base stations provide shared links for traffic demands. The positions of all 10 base stations are situated at specific positions. Hence, we complete the topology modeling. Next, we formulate the optimization algorithm for solving the indicated problem. Besides, We introduce two different coefficients to fulfill spectral efficiency and prioritization in the optimization algorithm which will be explained in the following chapters. Since the problem is regarded as linear programming problem, we use mainly use AMPL to verify the algorithm. Hence, we create 3 scenarios in MATLAB and implement them in AMPL. The final optimal solutions prove the correctness of the algorithm. The rate allocation assigned to each subscribers meets the requirement of Max-min fairness criterion. Furthermore, the spectral efficiency is taken into consideration, and prioritization is basically fulfilled.

In the second chapter, multiple network optimization techniques would be introduced. This chapter also includes the description of fair network. Additionally, a brief introduction of previous study is described. The third chapter is mainly about rate allocation in described heterogeneous cellular system, and network flow formulation. Also, the algorithm of the optimal solution is introduced in the third chapter. In the next chapter, which is chapter 4, provides the implementation details and numerical results of the optimal solutions. The last chapter summarize the whole thesis and make a conclusion, as well as a discussion about possible future improvements of the implementation and future research topics.

2. NETWORK OPTIMIZATION TECHNIQUES

In this chapter, an optimization technique is introduced and applied in the following chapters. This chapter is the background knowledge before introducing the principle of the implementation. Four sections are included in this chapter. In the section 2.1, the definition of network flow problem and its notation will be fully described. After that, we will introduce Fair networks and optimization techniques associated with fair demand volume allocation in section 2.2. In particular, we will further describe one famous assignment principle called Max-Min Fairness (MMF) in section 2.3. In the last section of this chapter, the previous study of this topic and its shortcomings are briefly introduced.

2.1 Network flow problem

Network flow problem (NFP) is a type of network optimization problem in order to route as many packets as possible on a given network. We will use a simple network which includes three nodes connected with each other as a illustration. The topology of this sample network looks like a triangle, as shown in Figure 2.1;

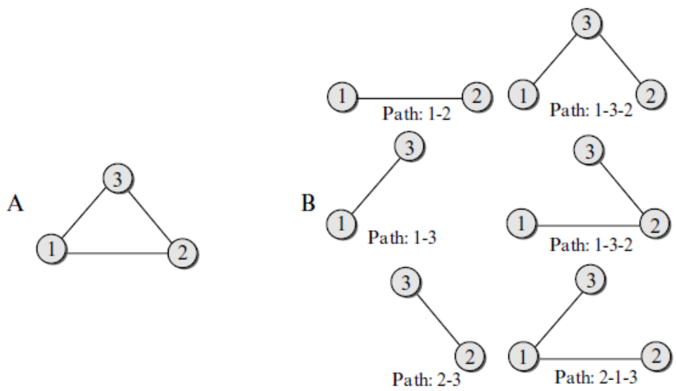

Figure 2.1: (A)Three-Node Network Example and (B) All Possible Paths for the Three-Node Example [9].

In this example, nodes can be routers in the Internet, telephone switches in the telephone network, or digital cross-connects in the SONET network[9]. In the following description, *node* is used to represent different types of routing or switching devices in a network. Besides, we use the name *demand volume* to identify either the traffic volume or the required bandwidth between a pair of nodes. For instance, a pair of nodes is named a *demand pair*, or just *demand.* In order to keep this example simple enough, we could assume that the links are undirected. Moreover, suppose that demand volumes between nodes 1 and 2, between nodes 1 and 3, and between 2 and 3 are 5,7,8, respectively. Here, h is used to stand for the demand volume:

$$h_{12} = 5, h_{13} = 7, h_{23} = 8.$$

Clearly, two paths which could be routed is available for the demand volume for a pair of node in this network.For instance, for the demand pair with end nodes 1 and 2, which is denoted as $< 1, 2 >$. The demand volume of this pair nodes could be either routed over the direct-link route 1-2 or the route 1-3-2 via node 3(Figure 2.1B). The number of the demand volume routed on each path is mainly depend on the network design objective. Therefore, if x is used to represent the unknown demand path-flow variables or simply flows, the expression for demand pair $< 1, 2 >$ could be wrote as:

$$x_{12} + x_{132} = 5(= h_{12}). \tag{2.1}$$

Note that the subscripts of variables x to denote the route or the path, here are paths 1-2 and 1-3-2. Likewise, we could have the equations for demand pairs <1,3> and <2,3> as follows:

$$x_{13} + x_{123} = 7(= h_{13}) \tag{2.2}$$
$$x_{23} + x_{213} = 8(= h_{23}). \tag{2.3}$$

Obviously, the path flows should be non-negative for all paths. Another item we should take into consideration is the *link capacity* or referred to as *link bandwidth.* For differentiating demand from links, the links is denoted by 1-2,1-3, and 2-3, while the capacity associated with these links is denoted by c_{12}, c_{13} and c_{23}, respectively. It should be emphasized that the demand volume could be between any pair of nodes when the link connects two nodes directly. Moreover, the unit of the demand volume has to be consistent with the unit of link capacities. Therefore, it could be fulfilled less hard for a link for which the capacity is given in terms of raw link rate, i.e. Mbps.

Then, we need to find out which flows might use different links. It is manifest that flow variables x_{12}, x_{123}, and x_{213} use link 1-2 which has a capacity of c_{12}. The basic requirement in network is that the link load cannot exceed the capacity of the link. Hence, we could have the following inequality for link 1-2:

$$x_{12} + x_{123} + x_{213} \leq c_{12} \tag{2.4}$$

Similarly, for other two links 1-3 and 2-3:

$$x_{13} + x_{132} + x_{213} \leq c_{13} \tag{2.5}$$
$$x_{23} + x_{132} + x_{123} \leq c_{23} \tag{2.6}$$

Assume that the capacity of the first two link is 10 and third on is 15; then, we could have:

$$c_{12} = c_{13} = 10, c_{23} = 15.$$

Based on what is discussed so far, it has following set of linear expressions where x are unknowns for all three demands considered.

$$\begin{aligned}
&x_{12} + x_{132} = 5\\
&x_{13} + x_{123} = 7\\
&x_{23} + x_{213} = 8\\
&x_{12} + x_{123} + x_{213} \leq c_{12}\\
&x_{13} + x_{132} + x_{213} \leq c_{13}\\
&x_{23} + x_{132} + x_{123} \leq c_{23}\\
&x_{12}, x_{132}, x_{13}, x_{123}, x_{23}, x_{213} \geq 0
\end{aligned} \tag{2.7a}$$

As a matter of fact, system(2.7a) has several solutions and defines the set of all feasible solutions. Here comes the question that which specific feasible solution is the best solution. To tackle this question, we need to realize what is essential as far as the goal of network design is concerned. In other words, we now need to know the objective function of the network design.

Let us assume the cost of routing one unit of flow on every link along its path is 1, and the we need to minimize the total routing cost. Therefore, the objective function could be wrote as:

$$F = x_{12} + 2x_{132} + x_{13} + 2x_{123} + x_{23} + 2x_{213} \tag{2.7b}$$

For example, the unit path cost on path 1-3-2 is 2 because the path 1-3-2 consists of two links where the unit link cost with respect to routing is 1.

In summary, the problem is to minimize the objective function (2.7b), which is subject to the constraints in system (2.7a). So, we could write the complete routing minimization problem discussed so far as problem 2.1:

Minimize:

$$F = x_{12} + 2x_{132} + x_{13} + 2x_{123} + x_{23} + 2x_{213}$$

Subject to:

$$\begin{aligned}
&x_{12} + x_{132} = 5\\
&x_{13} + x_{123} = 7\\
&x_{23} + x_{213} = 8\\
&x_{12} + x_{123} + x_{213} \leq c_{12}
\end{aligned} \tag{2.8}$$

$$x_{13} + x_{132} + x_{213} \leq c_{13}$$
$$x_{23} + x_{132} + x_{123} \leq c_{23}$$
$$x_{12}, x_{132}, x_{13}, x_{123}, x_{23}, x_{213} \geq 0$$

Problem 2.1 is a common example of a multi-commodity network flow problem. Since there are multiple demands, it is called multi-commodity. Those demands should be routed in the network simultaneously and complete for available resources (link capacities). It can also be regarded as linear programming problem because all the constrains and objective function are linear.

The final target is to find the optimal solution for problem 2.1. Basically, it is not difficult to solve this problem with some common sense. The final optimal solution for problem 2.1 is:

$$x^*_{12} = 5, x^*_{13} = 7, x^*_{23} = 8.$$

while other flow variables are 0,and the minimum routing cost F is 20. This solution is optimal and feasible since it meets all the constraints. Furthermore, the optimal solution is unique in this case. However, at most of time, it could have several optimal solutions.

We should realize that it could become more complex in other cases. For example, a small variation of above problem 2.1 can be made. Assume that the routing cost of a unit of flow is twice as expensive to go on the direct path compared to the alternate path. The objective function of this variance is:

$$F = 2x_{12} + x_{132} + 2x_{13} + x_{123} + 2x_{23} + x_{213} \tag{2.9}$$

Similar method could be use to solve this variance, and simple route all the traffic through alternate paths. Nevertheless, the capacity constraints will not be satisfied this time. But, there is still available solution for objective function. In this case, constraints is not changed but only the objective function. So, the optimal solution obtained in problem 2.1 is still a feasible solution but not the optimal one. It can be found that the optimal solution for the revised problem is somehow between the previous solution and the situation when all demand volumes are routed through the cheaper path. The cheaper path refers to the multiple-link path in the variance problem. Now, we could have the final results without explaining how to solve it:

$$x_{12} = 0, x_{13} = 1, x_{23} = 4, x_{132} = 5, x_{123} = 6, x_{213} = 4.$$

While the minimum routing cost F* is 25 in this case.

Based on the discussion above, we introduce the definition of network flow problem, and illustrate it in details. Finally, we could have two results from the problem 2.1 and its variance:

- The objective function could significantly affect the optimal solution to a problem, and method of solving it.
- It is necessary to completely understand and use the objective function for a particular network; otherwise, the optimal solution obtained may be totally meaningless.

2.2 Notions and notation in network flow problem

As can be found in section 2.1, a notation is used to denote nodes, links, demands, and path flows for a multi-commodity flow problem. This notation is not appropriate in the general case even though it works well in the three-node network. Therefore, we would introduce a new notation to replace it in this section. In the rest of this thesis, we will start to use this new notation to represent nodes, links, demands, and path flows.

In section 2.1, we have used a notation referred to as *node-identifier-based notation*. It means that all demands and paths are easy to follow from a node-reference point of view. However, this notation has three main shortcomings. First of all, not all node pairs have demand and/or some nodes are not directly connected. Then, paths may consists of multiple intermediate nodes. Lastly, flow variables may have indices of various length. Assume that we have 50 nodes in a network, the demand does not exist between node 5 and node 20. Imagine there is no any demand for some pairs of nodes, and we somehow need to indicate some specific pairs of nodes do not have any demand. Using the node-identifier-based notation, this should be listed clearly within the context of the model. For example, we have demand h_{ij} except h_{520} ($< i, j >=< 5, 20 >$) and so on. It finally results in deviating from the main flow of understanding a problem formulation and become a distraction. More importantly, there are growing number of possible paths between two nodes when the network becomes larger. The paths can also be of a variable number of links. It will cause troubles in representing multiple paths for certain demand pair when each path may go through various intermediate nodes.

To solve such problem, we introduce a new notation called link-demand-path-identifier-based notation . This notation is compact and allows to list only the compulsory objects. It is much more easy to capture, formulate and understand for multi-commodity network flow problems. It is also beneficial for making algebraic manipulations on the formulated problems.

Now, the principle of the link-demand-path-identifier is explained in details as follows. We would assign indices from 1 to the total number of demand pairs to those demand pairs that have non-zero demand volume. In this case, any pair of nodes without any demand is not listed at all. The three-node network shown in Figure 2.1 is still used to illustrate. We could write the demand pairs as follows:

demand pair $< 1, 2 > \longleftrightarrow$ label 1
demand pair $< 1, 3 > \longleftrightarrow$ label 2
demand pair $< 2, 3 > \longleftrightarrow$ label 3.

In general, notation D is used to represent the total number of demand pairs, and index d is used to label those demands. Obviously, D is an positive integer, while the value of d should be between 1 to the total number of links. In particular, D equals 3 and d equals 1,2,3, in this example. Similarly, we use the method to list all links of the problem as follows:

link pair $1 - 2 \longleftrightarrow$ label 1
link pair $1 - 3 \longleftrightarrow$ label 2
link pair $2 - 3 \longleftrightarrow$ label 3.

Besides, we could use notation E to denote the total number of actual links in the network, and index e to mark the links. Then, E should equal 3, and e equals 1,2,3 in this case. Note that if there are multiple demands or links between the same pair of nodes, they can be simply included into the list of demands or links, respectively.[9]

Based on the discussion above, we could have the following equivalence mapping for demand volumes and link capacities for the three-node network.

$h_{12} \longleftrightarrow h_1, h_{13} \longleftrightarrow h_2, h_{23} \longleftrightarrow h_3$
$c_{12} \longleftrightarrow c_1, c_{13} \longleftrightarrow c_2, c_{23} \longleftrightarrow c_3.$

After successfully mapping the demand pairs and the links to the new notation, we could discuss about the transformation for the paths. The demand pair identifier is used as the first subscript in a path-flow variable. And, the second subscript is used as the label for the path for that particular demand pair. For instance, Pd stands for the total number of candidate paths for demand d, while index p labels the paths. Hence, P2 = 3 means that demand <2,3> is identified by label 2. The two candidate paths, 1-2 and 1-3-2, are labeled with p=1,2, respectively. Consequently, we could map the flow variables to their equivalence as listed below:

$x_{12} \longleftrightarrow x_{11}, x_{132} \longleftrightarrow x_{12}$
$x_{13} \longleftrightarrow x_{21}, x_{123} \longleftrightarrow x_{22}$
$x_{23} \longleftrightarrow x_{31}, x_{213} \longleftrightarrow x_{32}$

Based the transformations above, we could re-write the problem 2.1 using new notation as shown below:

Minimize:

$$F = x_{11} + 2x_{12} + x_{21} + 2x_{22} + x_{31} + 2x_{32}$$

Subject to:

$$\begin{aligned}
&x_{11} + x_{12} = h_1 \\
&x_{21} + x_{22} = h_2 \\
&x_{31} + x_{32} = h_3 \\
&x_{11} + x_{22} + x_{32} \leq c_1 \\
&x_{12} + x_{21} + x_{32} \leq c_2 \\
&x_{12} + x_{22} + x_{31} \leq c_3 \\
&x_{11}, x_{12}, x_{21}, x_{22}, x_{31}, x_{32} \geq 0.
\end{aligned} \tag{2.10}$$

Finally, it should be emphasized that both formulations are for the same problem, even if two formulations use different notations.

2.3 Fair networks

In this section ,we will describe optimization methods associated with fair demand volume allocation. We will start from discussing the assignment principle named Max-Min Fairness (MMF). The final optimization goal is to realize high bandwidth utilization in terms of fairness.

Fair networks are user-centric metric to satisfy fairness of resource allocation between users. Obviously, the traffic should be elastic. The elasticity means that each demand can consume any aggregated bandwidth assigned to its individual path(s), perhaps within certain predefined limits[9]. For instance, demands generates elastic traffic in a network, and those traffic adapts to the varying bandwidth currently assigned to them. The main problem is how to assign bandwidth to the demand paths so that the capacities of links will not be exceeded. Also, we should assure that the actual aggregated bandwidth volumes assigned to demands are allocated in a fair way.

We should take two factors into consideration to meet this intuitive requirement. First one is the fairness of allocation of resources between subscribers. The other one is the network throughput maximization. However, fairness of resource allocation between subscribers is a user-centric metric while network throughput is network-centric metric. Usually, we need to make some compromises between these two metrics. There are two famous principles of fairness allocation, which are Max-Min Fairness (MMF) and Proportional Fairness (PF).

Max-min fairness is the most widely known fairness criterion, and aims at maximizing the minimum of bandwidth allocation $x_d,\ d \in \{1, 2, \ldots, D\}$ subject to capacity constraints of links and satisfying non-negativity of allocation. Suppose that $\vec{x} = (x_1, x_2, \ldots, x_D)$ is the allocation vector sorted in non-increasing order. Then, $\vec{x}$ provides max-min allocation if it is lexicographically maximal among all allocation vectors sorted in non-decreasing order. Therefore, the allocation is called max-min optimal if it is impossible for one to increase the allocation for some demand i at the expense of demands j with greater allocation.

On the other side, proportional fairness is a compromise between throughput maximization and MMF criterion. Normally, PF's goal is to maximize the sum of logarithms of the flows assigned to demands. Using basic mathematical knowledge, it could be easy to understand the reasonability of PF. At first, it avoids assigning zero volume to demands. Then, it is less beneficial to assign too much volume after considering the property of logarithm function.

Both topology model of a heterogeneous network and certain wireless systems have their constraints when choosing the fairness criterion. Firstly, the proportional fairness was developed based on wired networks, so PF is more beneficial for the short flows than longer flows when they both compete for resources. However, we could find that all the paths are of the same length in wireless heterogeneous networks. It means proportional fairness is equally beneficial with max-min fairness.

As a matter of fact, the final optimal solutions obtained with max-min fairness and proportional fairness are totally the same in our topologies. Beside, another observation is that max-min fairness focus on providing as fair allocation as possible, which is meaningful for wired networks but inappropriate for wireless networks. The current distance from the base station could dramatically has an effect on the choice of modulation and coding scheme (MCS) , and eventually affects the current effective rate provided by the network. This is the reason why max-min fairness is inappropriate for wireless networks. Before max-min implementation, those subscribers with small spectral efficiency need to be assigned with larger bandwidth allocations. It means that we could decrease the total throughput of the entire network when we try to provide exact fair allocation, which is unexpected for network operators.

In order to balance between fairness of allocation and network throughput, we introduce a simple bandwidth-based max-min fairness. We tend to the division of available bandwidth as fair as possible instead of trying to provide the fair allocation in terms of rates offered to the subscribers. Using this method, we could provide fair allocation with respect to the set of frequencies obtained form the network for a specific base station. Moreover, stations which are more close to their base stations could obtain higher rate than those situated further with this method. More importantly, this criterion also avoids infinitesimally small rates since all the subscribers are provided with the same set of frequencies. Note that the highest rate should not exceed the upper limit of the fastest MCS which is used in a particular technology. Therefore, the proposed scheme makes a compromise between network throughput benefiting from the dynamic nature of modulation and coding schemes and the fairness of rate allocations.

It is also worth emphasizing that proposed criterion may be considered as a variance of proportional fairness for cellular system, where the stations situated at larger distances from the base station are penalized more heavily. In the rest of this thesis, the optimization problem could still be classified with linear programming problems, even though the proposed objective function provides an analogue of proportional fairness allocation for wireless networks. Since it is a linear programming problem, we could use effective solutions algorithms, such as simplex or interior-point algorithms. The computational complexity of these algorithms scales well with the dimension of the problem.

To sum up, through comparing two fairness criteria, PF works better than MMF in terms of throughput. However, PF could support short flow, and lead to a less fair solution. On the contrary, MMF is relatively more fair at the expense of throughput. Whereas, the problem mentioned in this thesis, the MMF and PF have no difference when obtain the optimal solution. Thus, we choose MMF criterion to solve in view of mathematical complexity.

2.4 Previous study

In this section, the previous study of this topic will be briefly introduced. This thesis is the further extension of 'On the optimal assisted rate allocation in n-tier multi-RAT heterogeneous networks' by Qiao Wang.

In the previous study, the objective of it is to solve N-tier heterogeneous network resource allocation problem through considering it as a network problem.[10] The method he used is to formulate a heterogeneous problem as a network flow problem, it turns out to be an effective way for assisted network selection.

In the mentioned thesis, the author has introduced several parts. At the beginning, he introduced the fundamental and core notions and notations of network flow problem. Then, the concept of fair network and corresponding criterion is described in details. Next part is mainly about the formulation of the targeted 3-tier HetNet problem. After that, author proposed two available algorithm, which are based on NBT-1 and NBT-3, respectively. Before the implementation, author introduce the important tool used in his thesis, which is AMPL. Author gave a brief and specific introduction to AMPL language with an understandable example, and how the concentrated problem is established in AMPL. He also introduced the principle of solving the problem with CPLEX. In the implementation part, author illustrated with 5 topologies established with MATLAB. After the implementation, author analyzed the numerical results for those 5 topologies. In the final part, he made a comparison with two counterpart heuristic schemes, which are WiFi-preferred and Max-usage.

Eventually, the comparison of WiFi-preferred, proposed Max-min and Max-usage is shown as in Figure 2.2;

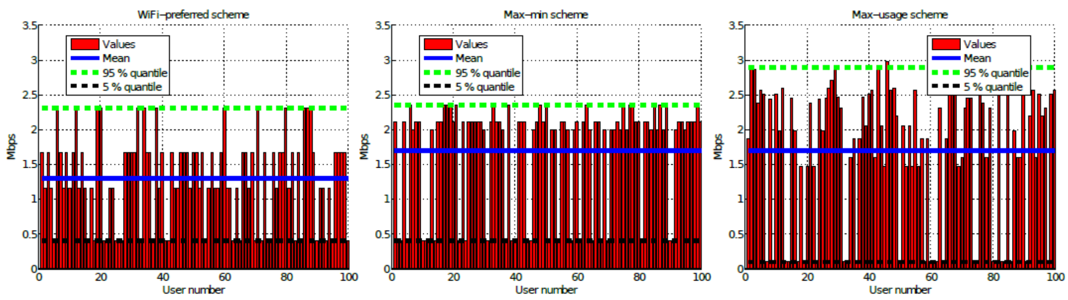

Figure 2.2: Per-user throughput performance comparison: (a) WiFi-preferred, (b) Proposed Max-min, and (c) Max-usage schemes [6].

Figure 2.2 show the results of the comparison of performance with throughput in the mean, 5%-quantile and 95%-quantile[10]. The conclusion is that proposed scheme Max-min algorithm in subfigure(b) is higher than that of WiFi-preferred scheme with respect to average per-user throughput. Even though the second and the third schemes are sharing the same value of "mean" quantile, the throughput of the subscribers served by merely macro-LTE is significantly improved. Thus, Max-min algorithm not only combines the benefits but also avoids shortcomings of both counterpart heuristic schemes[6].

Besides, author has following findings. The first one is that MMF is the most fair way for resource or bandwidth allocation among subscribers. In particular, the traffic demands are set to have one or more paths for traffic realizing in his implementation process. The proposed algorithm has correctly solve the fair allocation between subscribers in terms of MMF criterion. Then, author highly recommend AMPL modeling software as an excellent tool for modeling network problem, especially the problem is a typical linear programming problem. AMPL associated with CPLEX solver turns out to be very ideal for solve dual variables problem. It is not only a ideal tool but having high efficiency when solving the problem. It usually costs a few seconds for complex computation. The last finding is that MMF is an excellent scheme for the typical heterogeneous networks at present.

Even if the results prove to be totally correct in previous study, it still has a few shortcomings listed as follows:

- The coverage of macro-LTE in previous study is not reasonable in terms of radius, since there are only 20 subscribers within the coverage.
- None subscriber could be inside of any coverage of micro-LTE or Wi-Fi base stations. Then, the topology is meaningless to discuss in this case.
- There is no modulation and coding schemes, since it is necessary during wireless transmission.

- In the previous study, the prioritization is not introduced for further improvement.

Based on the description above, we would further extend this topic in two aspects, which are spectral efficiency and priority in our thesis. In the next chapter, an improved algorithm is introduced and discussed. Furthermore, multiple more realistic scenarios would be created to illustrate the implementation process. At the same time, the author of the previous study should be fully appreciated for his time and effort on this topic. With the results of the previous study, we could introduce a more realistic and sophisticated extension in this thesis.

3. RATE ALLOCATION IN HETEROGENEOUS SYSTEM

This chapter discusses the most important theoretical part, and has 5 sections for detailed description of rate allocation and formulation. In section 3.1, the detailed description of particular heterogeneous system and resource allocation will be shown. The topology modeling part will be illustrated in the section 3.2. In the next section, the network flow formulation will be further expressed and explained. After that, optimal solution algorithm will be presented in section 3.4. Last section will mainly discuss about how to apply this model to prioritization.

3.1 Description of heterogeneous system

LTE stands for Long Term Evolution and it was started as a project in 2004 by telecommunication body known as Third Generation Partnership Project (3GPP) [12]. It provides for a downlink rate of up to 150 Mbps and an uplink rate of up to 50 Mbps. The main objective of LTE is to increase downlink and uplink maximum data rates, to provide low latency, to improve cell edge performance(in terms of bit rate), and to improve spectral efficiency. Mobility and seamless handoff were requirements from the start, as was a requirement for central management of all nodes[13].

On one hand, macrocell provides the main radio coverage infrastructure for a mobile network[14]. The antennas are placed at a height which has a clear view over the surrounding buildings and terrain. The coverage of macro network depends on frequency capacity and clutter. Usually, the radius of coverage is 1 to 10 km. Meanwhile, microcells provide infill radio coverage and additional capacity where there are high numbers of users within urban and suburban macrocells[14]. The height of microcell's antenna is lower than the height of macrocell's antenna. The coverage of microcells is smaller than 1 km. Also, microcells are embedded within macrocells. They are classified as one type of small cells because they are small compared to macrocells.

The differences between macrocells and microcells are the mobility and desired data rates. Macrocells support high speed moving subscribers while microcells support slow moving subscribers. Moreover, macrocells meet the requirement of stream media service while microcells only provide data service.

On the other side, heterogeneous cellular networks are expected to become more varied. It means that the macro network will be improved by layers of technology, including Wi-Fi and small cells. Hence, it will offer carriers the flexibility to more efficiently provide desired coverage and capacity. Those Wi-Fi and small cells could also offload some of the traffic from macro base stations. Wi-Fi is a trademark of the Wi-Fi Alliance, which is an industry association promoting the standardization and interoperability of wireless local network (WLAN) connectivity based on the IEEE 802.11 series of standards[15].

Wi-Fi makes devices connect and access to the Internet possible. The connection could be direct or through a router without any physical association with a wired network. The 802.11 standard defines several versions of WLAN connectivity, and continues to evolve as needs advance and technology evolves[15]. There are several available options at present. Firstly, 802.11a was the first standard aimed at enterprise-class wireless LAN technology, and could offer many advantages over previous options. The 5 GHz band that 802.11a operates in is not highly populated, so there is less congestion to cause interference or signal contention[16]. Then, the IEEE established 802.11b in 1999 to improve the data rate of the original 802.11 standard-defining rates up to 11 Mbps[17]. However, 802.11b operates in the 2.4 GHz, which cause interference due to many other devices operating in the same frequency.

After that, another standard named 802.11g was proposed by IEEE. It also operates in the 2.4 GHz. In addition, it includes the same OFDM based transmission scheme as 802.11a. The peak throughput(physical layer) of 802.11g reaches 54 Mbps, except the forward error correction codes[18]. 802.11n further improves on the previous 802.11 standards by adding multiple-input multiple-output antennas. As a consequence, the throughput could be up to 600 Mbps over the air estimates[19]. The most recent standard is 802.11ac, which is designed to use the 5 GHz spectrum. Since the 2.4 GHz spectrum is congested with current Wi-Fi devices, the 5 GHz spectrum could have less interference and congestion[20].

In order to provide services to a explosive growth population, we need to deploy more base stations. The spectrum where macro-LTE base station works on is 2 GHz, while the spectrum where Wi-Fi base station works on is 2.4 GHz. Hence, we deploy 10 Wi-Fi base stations and 1 macro-LTE in the same topology to meet the requirements of subscribers. Furthermore, macro-LTE will not have interference with Wi-Fi base stations. However, there is still minimal interference between Wi-Fi base stations.

3.2 Topology modeling

The heterogeneous system has N subscribers and two BS tiers (called here layers). Layer 1 (LTE macro-cell) has N_1 BSs, while Layer 2 (Wi-Fi known as access points)has N_2 BSs Hence, heterogeneous system has $M = N_1 + N_2$ BSs in total. Besides, one physical aggregator which assembles traffic after the subscribers splits it over several RATs. On the other side, the modulation schemes used in the system are QPSK, 16-QAM and 64-QAM, while the coding scheme is Forward Error Correcting(FEC) coding. Considering the modulation, QPSK carries 2 bits per symbol, 16-QAM carries 4 bits per symbol and 64-QAM carries 6 bits per symbol. The overall coding rates of both QPSK and16-QAM are 1/2 and 3/4, and the overall coding rates of 64-QAM are 2/3 and 3/4. Thus, the spectral efficiency of QPSK, 16-QAM and 64-QAM is as tabled below.

MSC	Bits per symbol	Overall coding rate	Efficiency (bit/s/Hz)
QPSK	2	1/2	1
QPSK	2	3/4	1.5
16-QAM	4	1/2	2
16-QAM	4	3/4	3
64-QAM	6	2/3	4
64-QAM	6	3/4	4.5

Table 3.1: The MCS of QPSK,16-QAM and 64-QAM.

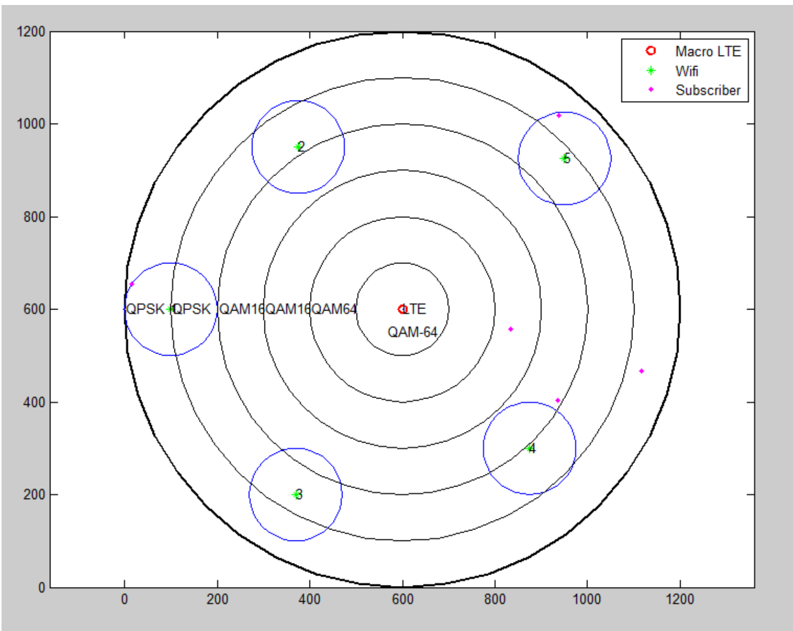

Figure 3.1: A simple example of the system

As shown in the Figure 3.1, a sample system has 5 subscribers, 1 macro-LTE BS, 5 Wi-Fi BSs and 5 sub circles. These 5 sub circles divide the coverage area into 6 sub areas to provide different spectral efficiencies. One macro-LTE BS is located in the center of the coverage area, and 5 IEEE based wireless local area networks(WLAN) surround the macro-LTE BS. Furthermore, 10 subscribers are uniform distributed randomly in the coverage area. Meanwhile, 2 subscribers also have access to wifi BSs when all of the subscribers have access to macro-LTE BS. Assuming subscriber 1 and subscriber 5 are in the coverage area of wifi 1 and wifi 5 respectively.

In order to justify the further discussion of the topology modeling, we make the following assumptions:

- A subscriber may access to on or multiple BSs at each layer;
- Traffic generated by a subscriber is greedy(full-buffer) and elastic;
- Locations of the wifi BSs are known;
- Rate obtained at layers depend on the distance to the BSs and spectral efficiency;
- No interference affects the performance.

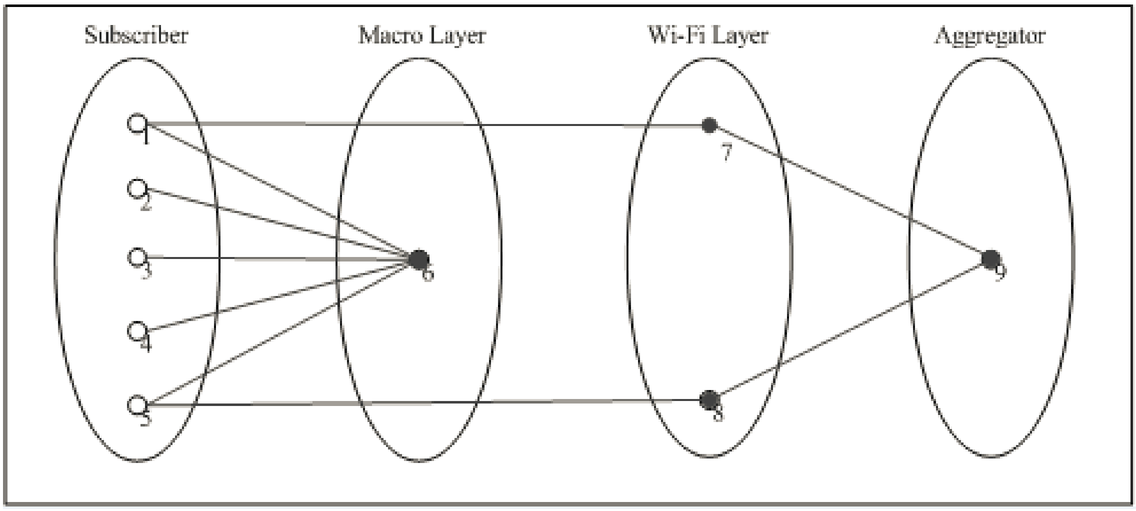

Figure 3.2: A sample topology of five-user system

As shown in Figure 3.2,and since links between the subscribers and BSs are shared, defining the link rates becomes impossible. In the rest of this thesis, only bifurcated resource allocation will be considered. Assuming a subscriber may simultaneously employ more than one wireless interface, so there is an aggregator (node

9 in Figure 3.2) terminating virtual tunnels over different wireless access networks. It is obvious that the topology illustrated in Figure 3.2 is redundant since the links connecting the BSs to the aggregator should have equal or higher capacity than the ones provided by these BSs at the air interface.

To simplify this topology, some redundant links which bring no additional constrains will be removed. Here node 1 is the logical aggregator, while node 2 acts as the physical aggregator (corresponds to node 9 in Figure 3.2). The number of links in such a system equals the number of BSs, that is, M=N1+N2. At the same time, the capacities of links are equal to the capacities of corresponding BSs. Hence, the simplified final modeling topology is illustrated in Figure 3.3.

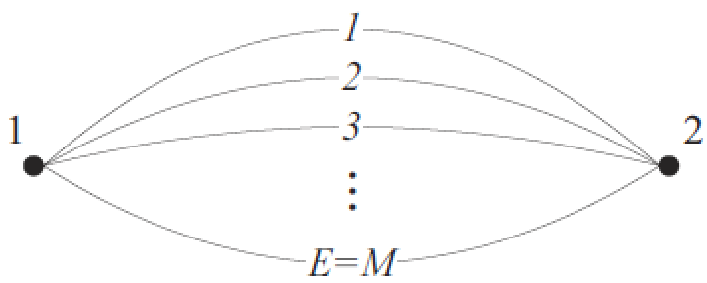

Figure 3.3: Final topology of five-user system

3.3 Network flow formulation

In the earlier section, only two nodes in the simplified final topology, the logical node and the physical aggregator. These two nodes are connected by M=N1+N2 links, where Ni represents the number of the BSs at layer i associating at least one subscriber. Taking the resource provided by the system into consideration, that is the bandwidth is available at the corresponding air interfaces.

The first step of solving the network flow formulation is to determine the resource allocation which is needed to assigned to demands. For example, a certain fairness criteria should be applied, and mentioned modulation and coding schemes (MCS) above need to be considered. Thus, we could convert the frequency allocation to the actual rate provided. Recall that only the bifurcated solution is considered in this thesis. First of all, the problem is defined for the max-min allocations, then is reduced to the specific criterion.Making N to be the number of demands, and it could be realized between two nodes marked with d,d=1,2,. . . ,N. Then, the demand volumes (in Hz) are expressed as

$$h_d,\ d = 1, 2, \ldots, N. \tag{3.1}$$

and the demand volumes are unknown in advance if greedy elastic traffic is assumed. We should determine the values of the demand volumes(hd), so the specific fairness criterion could be satisfied. Assume that,

$$\Pi_d,\ d = P_{d1}, P_{d2}, \ldots, P_{dPd} \tag{3.2}$$

where P_d is the number of available paths for each demand, such as the number of the BSs a subscriber is associated with. Since all the subsets $P_{dp},\ p = 1, 2, \ldots, P_d$, d=1,2,...,N, have only one link connecting the logical node and the physical aggregator, these paths are readily available in the final topology. More specifically, demand 5 is in the coverage of macro-LTE and Wi-Fi station, so it has two paths to realize its demand requirement, which are 1 and 2, representing macro-LTE and wifi station respectively. Hence, the paths set of it is $\Pi_5 = \{P_{51}, P_{52}\}$, where P_{41}=\{1\},P_{42}=\{2\}.Even though, merely using paths as additional variables to formulate the task is not a very ideal approach, in order to avoid confusion, we still use them to distinguish the classic network flow problem notation.

The flow of demand(the amount of flow, in Hz) d assigned to the path p is denoted as

$$x_{dp},\ d = 1, 2, \ldots, N.\ p = 1, 2, \ldots, p_d, \tag{3.3}$$

Furthermore, the current spectral efficiency of the demand d is denoted as

$$s_{dp}(V),\ v = 1, 2, \ldots, V. \tag{3.4}$$

where v is the currently used modulation and coding scheme (MCS) and $sdp$is realized over the path p measured in bits/s/Hz.

It is noteworthy that the current spectral efficiency is determined by the MCS and propagation conditions of between a mobile node and BS. Recall the section 3.1, the detailed spectral efficiencies used in the current topology could be found in table 3.1.

Obviously, spectral efficiency sdp represents how efficient the flow allocation $x_{dp}$is currently used. Based on the max-min fairness criterion, more bandwidth should be assigned to the mobile nodes which have worse spectral efficiency. The overall rate provided to demand d is computed as $h_d$if the flow allocation x_{dp} could be changed to the actual bitrate obtained over p as xdpsdp. Then,

$$h = \textstyle\sum_{p=1}^{P_d} s_{dp} x_{dp},\ d = 1, 2, \ldots, N. \tag{3.5}$$

Meanwhile, the link-path-incidence variables δ_{edp} is introduced

$$\delta_{edp} = \begin{cases} 1, & e \bigcap P_{dp} = e \\ 0, & otherwise \end{cases} \tag{3.6}$$

If the path p of the demand d uses the link e, then variable δ_{edp} equals 1. After that, the capacity constrains could be expressed as

$$\sum_{d=1}^{N} \sum_{p=1}^{P_d} \delta_{edp} x_{dp} = c_e, e \in \{1, 2, \ldots, E\} \tag{3.7}$$

where c_e is the capacity of links between logical aggregator and physical aggregator. Therefore, the optimal rate assigned to demand could be obtained using the constrains above. Besides, max-min fairness criterion would be satisfied when taking the spectral efficiencies into account. In this case, MCSs currently used at the mobile node determine what the rate is provided in the system.

The constrains above also give another implication that no links in the network will be overloaded by the demands using these links. Since the problem discussed here could be converted to fair network capacitated problem with varying capacity, we could replace all the inequalities with equalities in the constrains above.

As far as we know, the objective of max-min fairness criterion in heterogeneous system is to find a vector $\vec{x}$ which are lexicographically maximal. To accomplish this objective, we need to lexicographically maximize the demand volume variable vector $\vec{h} = \{\vec{h_1}, \vec{h_2}, \ldots, \vec{h_N}\}$. The demand volume variable vector should be sorted in non-decreasing order, and the flow of demand x should be continuous and non-negative.

In terms of optimization problem, this problem could also be classified as a typical linear programming problem. That is the reason why we choose a modeling language named AMPL for solving the problem.

On the basis of the above discussion, the final formulation of the problem is as following:

indices $d = 1, 2, \ldots, N$
$p = 1, 2, \ldots, P_d$
$e = 1, 2, \ldots, E;$

constants $\delta_{edp} = 1$ if e belongs to the path p of demand d;
otherwise, it equals to 0.
c_e is the link capacity.

variables x_{dp} is the flow assigned to path p of demand d, x_{dp} is continuous and non-negative.
h_d is the total bandwidth assigned to demand d
s_{dp} is the current spectral efficiency of demand d
r_d is the overall rate provided to demand d

objective to lexicographically maximize the demand volume variable vector $\vec{h} = \{\vec{h_1}, \vec{h_2}, \ldots, \vec{h_N}\}$ sorted in non-decreasing order.

constrains $\sum_{p=1}^{P_d} s_{dp} x_{dp} = h_d, d = 1, 2, \ldots, N.$
$\sum_{d=1}^{N} \sum_{p=1}^{P_d} \delta_{edp} x_{dp} = c_e, e \in \{1, 2, \ldots, E\}$
all $x_{dp} \geq 0$

3.4 Optimal solution algorithm

In this section, a optimal solution algorithm will be proposed to solve our problem mentioned above. Since the 2-layer cellular network problem has been formulated in section 3.2, the solution algorithm could base on the flow formulation. In chapter 2, a solution algorithm of optimal assisted rate allocation in random positions within the coverage has been explained. The further extension of this solution is going to applied in this thesis to introduce the spectral efficiency. As a consequence, the solution algorithm will be substantially more complicated compared with previous algorithm. It actually an extended version of the previous algorithm.

For another, the problem is classified as linear programming problem which consists of basic water-filling algorithm. The water-filling algorithm mainly addresses the maximum allocation that could be allocated to all the flows simultaneously and further refines of allocations of specific flows. On the contrary, the first stage may have multiple solutions compared with the unique solution of the single path problem. The linear programming problem(3.3.1) is as follows:

Maximize: Δ,

Subject to:

$$\sum_{p=1}^{P_d} s_{dp} x_{dp} = h_d, d = 1, 2, \ldots, N. \tag{3.8}$$

$$\Delta - h_d \leq 0 \tag{3.9}$$

$$\sum_{d=1}^{N} \sum_{p=1}^{P_d} \delta_{edp} x_{dp} = c_e, e \in \{1, 2, \ldots, E\} \tag{3.10}$$

$$\text{all } x_{dp} \geq 0, d \in \{1, 2, \ldots, N\}, p \in \{1, 2, \ldots, p_d\} \tag{3.11}$$

The algorithm proposed is based on the non-blocking test, aka NBT for checking whether there is still h_d which could be increased. NBT highly determine the efficiency of algorithm. More precisely, in rest of the section, two algorithms will be compared with NBT. First algorithm with NBT1 is less efficient and more complex than the second algorithm with NBT3. Nevertheless, the first algorithm is still worth introducing since it is the fundamental principle for solving capacitated flow problem. Meanwhile, the second algorithm with NBT3 is based on dual variables of specific constrains. The first algorithm is also used in the implementation procedure for more convenient and easy to be implemented.

Two feasible algorithms(algorithm 3.3.1 and algorithm 3.3.2 will be explained and compared in detail as following.

Algorithm 3.3.1

Step 1: Let(Δ^*, x^*, h^*) be the optimal solution of LP (3.3.1). Set $Z_0 := \emptyset$, $Z_1 := \{1, 2, \ldots, N\}$, and $\Delta_d := \Delta^*$ for each $d \in Z_1$.

Step 2: Perform NBT1: start considering each $d \in Z_1$ to check whether the total allocation h_d could be made greater than Δ^* without decreasing the already found maximal allocation of other demands d'. NBT is performed during the check. If there is no blocking demand in Z_1,then move to step 3. In other words, demand $d \in Z_1$ is blocking when h_d could not be increased, the check of NBT move to the next step. Otherwise, it means that the demand d is found when the first blocking demand. Under this situation, we should add d to set Z_0 and remove d from Z_1.This process will continue until $Z_1 = \emptyset$, then solution is the optimal solution(allocation vector) of the problem. Otherwise, the process will proceed to step 3. Note that Z_0 is basically the set of blocking demands.

Step 3: Solve the following linear programming problem:

Maximize: Δ
Subject to:

$$\sum_{p=1}^{P_d} s_{dp} x_{dp} = h_d, d \in \{1, 2, \ldots, N\}. \tag{3.12}$$

$$\Delta - h_d \leq 0, d \in Z_1 \tag{3.13}$$

$$\Delta_d - h_d \leq 0, d \in Z_0 \tag{3.14}$$

$$\sum_{d=1}^{N} \sum_{p=1}^{P_d} \delta_{edp} x_{dp} \leq c_e, e \in \{1, 2, \ldots, E\} \tag{3.15}$$

$$\text{all } x_{dp} \geq 0, d \in \{1, 2, \ldots, N\}, p \in \{1, 2, \ldots, p_d\} \tag{3.16}$$

Step 4: Then, (Δ^*, x^*, h^*) is used to solve the following linear programming problem for each $d \in Z_1$, and go back to step 2.

Recall that NBT1 is used to solve the flowing linear programming problem for each $d \in Z_1$:

Maximize: h_d
Subject to:

$\sum_{p=1}^{P_d} s_{d'p} x_{d'p} = h'_d, d' \in \{1, 2, \ldots, N\}$

$\Delta_{d'} - h_{d'} \leq 0, d' \in \{1, 2, \ldots, N\}$

$\sum_{d'} \sum_p \delta_{ed'p} x_{ed'p} \leq c_e, e \in \{1, 2, \ldots, E\}$

$all x_{d'p} \geq 0.$

The outcome of NBT1 is positive, and it means that demand d is not blocking if the optimal solution h_d is strictly greater than Δ'. Otherwise, it means that the considered demand d is blocking and the allocation of demand d could not be increased anymore. Obviously, the results obtained from step 1 and step are used by the calculation.

Based on the description above, it is apparent that algorithm is of polynomial complexity. Nonetheless, it could be very time-consuming and inefficient in terms of large networks with thousands of demands since it needs to solve large amounts of NBT1s. Hence, an improved and more efficient NBT is needed. It is commonly known that the most efficient test is NBT3, and NBT3 is based on dual variables. The detailed description of NBT3 is as following.

Let the vector $\vec{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_N)$ be the optimal dual variables of the equation '$\Delta - h_d \leq 0$'. And the vector $\vec{\gamma}$ will be used in efficient NBTs for step 2 of algorithm 3.3.1, referred to as algorithm 3.3.2.

According to the dual theory, the corresponding demand d is blocking if the dual variables of problem (3.3.1) $\gamma_d > 0$. In other words, the total bandwidth allocation h_d of demand d could not be increased further. In summary, d is blocking if $\gamma_d > 0$, i.e., $d \in Z_0$ after the first execution of step 2. Unfortunately, it does not imply that d is non-blocking when $\gamma_d = 0$.

Algorithm 3.3.2

Step 1: Set $Z_0 := \emptyset, Z_1 := \{1, 2, \ldots, N\}, \Delta_d := 0$ for all demands

Step 2: Solve the linear problem below:

Maximize: Δ
Subject to:

$$\sum_{p=1}^{P_d} s_{dp} x_{dp} = h_d, d \in \{1, 2, \ldots, N\}. \quad (3.17)$$

$$\Delta - h_d \leq 0, d \in Z_1 \quad (3.18)$$

$$\Delta_d - h_d \leq 0, d \in Z_0 \quad (3.19)$$

$$\sum_d \sum_p \delta_{edp} x_{dp} \leq c_e, e \in \{1, 2, \ldots, E\} \quad (3.20)$$

$$\text{all } x_{dp} \geq 0, d \in \{1, 2, \ldots, N\}, p \in \{1, 2, \ldots, p_d\} \quad (3.21)$$

Let Δ^* be the optimal value and γ_d be the dual variables of (3.18), respectively.

Step 3: First, put $\Delta_d = \Delta^*$ for each $d \in Z_1$. Then, let $Z_0 := Z \bigcup \{d \in Z_1 : \gamma_d > 0\}$ and $Z_1 := \{d \in Z_1 : \gamma_d = 0\}$. If set Z_1 equals 0, then stop. The vector $\vec{h} = (h_1, h_2, \ldots, h_d) = (\Delta_1, \Delta_2, \ldots, \Delta_d)$ becomes final optimal solution of the problem; otherwise return to Step 2.

Note that in algorithm 3.3.2, if the optimal Δ^* in step 2 is strictly greater than the optimal solution of the problem obtained in the previous iteration, then all demands d belonging to set Z_1 are non-blocking. Otherwise, at least one demand d in set Z_1 is blocking (it is not "true" non-blocking demand), and Δ^* cannot increased. Another significant observation is that there will be at least one with $\gamma_d > 0$ among the newly obtained optimal dual variable γ_d. Therefore, one or more blocking demands will be found. It is based on the property of dual variables γ_d below:

$\sum_{d \in Z_1} \lambda_d^* = 1$ and $\lambda_d^* \geq 0$ for $d \in Z_1$.

On the basis of the above description, Algorithm 3.3.2 is clearly more simplified compared with Algorithm 3.3.1 owing to using dual variables. It also proves to be more efficient and less time-consuming when using algorithm 3.3.2 instead of algorithm 3.3.1. Especially, only one linear programming problem is solved in AMPL coding process.

3.5 Application to priority allocation

In terms of priority allocation, the coefficient introduced would have a significant influence on the priority assigned to the specific subscriber. The main purpose of this section is to apply mathematical model to prioritization.

Recall the constrain in algorithm 3.3.2, which is $\sum_{p=1}^{P_d} s_{dp}x_{dp} = h_d, d \in \{1, 2, \ldots, N\}$. Different spectral efficiencies of the subscribers could be fulfilled by assigning different coefficients. Likewise, we could also use another coefficient to represent priority.

More precisely, we could assume p_{dp} to be the priority of path p of demand d. And, we could assign high priority with bigger value of p_{dp} to some subscribers who pay much more than ordinary subscribers. In algorithm3.3.2, the coefficient s_{dp} would be replaced by p_{dp} to fulfill prioritization. The application of prioritization has a bigger influence on the market than the one of spectral efficiency.Furthermore, a pricing scheme which is more customer-oriented would be made to attract customers. Some customers would need high rate internet while others only need basic internet rate. Hence, the priority could be used to divide various needs for internet rate. The internet service provider could also make higher profits than before.

Since, the simulation in the chapter 4 could prove the practicable of the spectral efficiency, then the priority could be fulfill by introducing another coefficient. Similar simulation could be made using AMPL and MATLAB. Even so, we need to fulfill spectral efficiency first, because modulation and coding during transmission are compulsory. After that, we could prove that coefficients could be introduced in the algorithm 3.3.2.

As a consequence, to attract market's attention, prioritization should be more focused since it will create higher profits.Spectral efficiency is necessary, while priority has extensive application prospect. In the next chapter, we will create three scenarios to test the algorithm. Nevertheless, only the spectral efficiency is considered. Likewise, we could use the same method to test the priority.

4. NUMERICAL RESULTS

After presenting the theoretical background and implementation, numerical results of the implementation will be introduced in this chapter. MATLAB and AMPL are the software used to generate 3 scenarios to prove the reasonability of the previous algorithm.

4.1 Introduction to AMPL

AMPL, an acronym for "A Mathematical Programming Language", is an algebraic modeling language for describing and solving high-complexity problems for large-scale mathematical computation (i.e. large-scale optimization and scheduling-type problems)[11].

Since AMPL is a modeling language, it can be used to describe optimization data, variables, objectives, and constraints. Meanwhile, AMPL has the similar syntax with mathematical expressions, so it make users master AMPL easily. The reason why we choose AMPL to simulate the algorithm is the problem is a linear optimization problem, and AMPL is a very ideal for solving linear problem.

There are three kinds of files in AMPL, which are .data file, .mod file, and .run file. First of all, .data file is used to create new parameters. And, .mod file is used to describe the objective function and constraints. While, .run file is used to load the model file and data file to solve the problem. A simple topology will be explained when illustrating the AMPL.

In the sample topology, there are one macro-LTE base station, one micro base station , and four subscribers. The link capacities of the macro-LTE base station and micro base station are 100 bps and 50 bps, respectively. In the same time, the coverage of macro-LTE base station overlaps the coverage of micro base station. Subscriber 1 is in the coverage of macro-LTE, and subscriber 2 is in the coverage of micro base state. Meanwhile, subscriber 3 and 4 are in the overlapping part. In other words, both subscriber 1 and 2 have only one path connecting to the base station, while subscriber 3 and 4 both have two paths connecting two base stations.As demonstrated in the Figure 4.1.

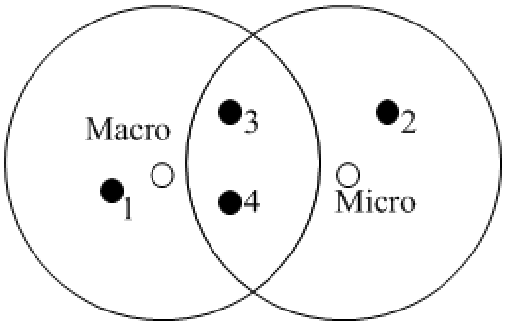

Figure 4.1: Sample topology

Therefore, we could create a data file in AMPL as in the figure 4.1.

```
data;
# 1 macro 1 micro
# 4 subscribers

param D := 4;
param E := 2;

param: link_capacity :=
1          100
2          50 ;

set Z0 := ;
set Z1 := ;
set Z2 := 1 2 3 4;

param: t:=
1       0
2       0
3       0
4       0 ;

set path_nos[1]:=1;
set path_nos[2]:=1;
set path_nos[3]:=1 2;
set path_nos[4]:=1 2;

set Path[1,1] := 1;
set Path[2,1] := 1;
set Path[3,1] := 1; set Path[3,2] := 2;
set Path[4,1] := 1; set Path[4,2] := 2;
```

Figure 4.2: Data file of sample topology in AMPL

Furthermore, model file and run file also needed to be introduced. In our problem, run file needs to load model file and data file to achieve the optimal solution. The run file is created as in Figure 4.2.

```
model topology2.mod;
data data6.dat;
option solver cplex;

repeat {
solve;
let {d in Z1} t[d]:=r;
for {d in Z1}
 { if second_con.dual[d] =0 then let Z2:=Z2 union {d};
    }

for {d in Z1}
   {
    if second_con.dual[d] >0 then let Z0:=Z0 union {d};
    }

 let Z1:=Z2;
 let Z2:={};
 }   until card(Z1)=0;

 if card(Z1)=0
        then {  printf "The optimal solution is below: \n";
                let {d in demand_nos} h[d]:=t[d];}

  #option display_1col
  display h;
  display x;
```

Figure 4.3: Command script file in AMPL

In Figure 4.2, the command option solver cplex is used to change the default solver MINOS to CPLEX, because CPLEX is more efficient and accurate compared with MINOS in solving linear and convex quadratic problems. The loop repeat$\{\dots\}$ until card$\{Z_1\}$ implements the main circular computations in the algorithm. First of all, the problem is solved by command *solve*. In order to solve the linear problem, the optimal solution Δ^* is obtained, which is represented as r. Assigning this value of r to each $t[d]$ for each $d \in Z_1$. This could be done by doing *let* d in $Z_1 t[d] := r$ where *let* is an assignment command.

After that, the program would check whether the set Z_1 is empty for two *for* loops are used to distinguish blocking demands from non-blocking demands according to the dual variables of second constraint in the algorithm. Therefore, the statement within the brackets could be executed for each d in set Z_1. The dual variable is obtained through $second_con.dual[d]$. The operator union$\{\}$ combines two sets into a new set. Then, the two statements *let* $Z_1 := Z_2$ and *let* $Z_2 := \{\}$, meaning Z_2 is an intermediate and transition set. Z_2 does not have any actual meaning and temporarily contains the non-blocking demands. Before next execution, set Z_2 would become an empty set and reused for transition again. Basically, the *for* and *set* statement distinguish two different kinds of demands.

Next, r is not using the command *until card(Z_1)* $= 0$. More precisely, the function card would compute the number of members in the set. If the set Z_1 is empty, the card(Z_1) would equals 0. Otherwise, the program would executed again to get the new optimal solution of r, and repeat the above process to distinguish more blocking demands from non-blocking demands until set Z_1 is empty. It means that there is no more non-blocking demands when set Z_1 is empty. In other words, the final optimal solution is found when the total demands allocation could not be increased further.

Next, we would explain the model file. And, the model file is created as in Figure 4.3.

```
param D > 0 integer;
param E > 0 integer;

set link_nos := 1..E;
set demand_nos := 1..D;
set Z1;
set Z2;
set Z0;

set path_nos{d in demand_nos};
param t{demand_nos};
param link_capacity {link_nos} >=0 integer;
param S{demand_nos,link_nos} >=0 ;
set Path {d in demand_nos,p in path_nos[d]} within link_nos;

param delta {e in link_nos,d in demand_nos,p in path_nos[d]}
    = if e in Path[d,p] then 1 else 0;

var r >= 0;
var h {demand_nos}>=0;
var x {d in demand_nos, p in path_nos[d]} >= 0;


maximize opti: r;

subject to first_con {d in demand_nos}:
    sum {p in path_nos[d]} S[d,p]*x[d,p] = h[d];

subject to second_con {d in Z1}:
    r - h[d] <=0;

subject to third_con {d in Z0}:
    t[d]-h[d]<=0;

subject to fourth_con {e in link_nos}:
    sum{d in demand_nos} (sum {p in path_nos[d]} (delta[e,d,p]*x[d,p]))
            - link_capacity[e]<=0;
```

Figure 4.4: Model file in AMPL

As shown in Figure 4.3, three sets have been declared, which are Z_1, Z_2 and Z_0. While Z_2 is an intermediate set for assignment without any actual meaning. We represent the collection of demands $\{1, 2, \ldots, N\}$ and the collection of links $\{1, 2, \ldots, E\}$ with *set demand_nos* $:= 1 \ldots D$ and *set link_nos* $:= 1 \ldots E$. Parameter D and E are positive integers. And, *param t{demand_nos}* means that parameter t has subscripts indexed from set *demand_nos*. Similarly with *link_capacity* and S when *link_capacity* means c_e and S means s_{dp} in mathematical mode.

More importantly, *path_nos{d in demand_nos}* means that for each demand d of *demand_nos* there is to be a set *path_nos*$[d]$. In other words, it indicates the total number of paths of each demand d. Since each demand has multiple candidate paths, this method is needed to distinguish the difference. Furthermore, for each demand d of *demand_nos* and p of *path_nos*$[d]$, it needs a set Path[d,p], which is a subset of *link_nos*. The set is significant since the value of the parameter delta greatly depends on it. Similar with declaring delta(δ_{edp}), we also set *param delta{e in link_nos, d in demand_nos, p in path_nos[d]}*. Besides, we replace Δ with r, and Δ_d with *t_d* for convenience.

Based on the illustration, we could have a preliminary acquaintance of AMPL. In the following sections, we would describe three different scenarios when simulating the algorithm in AMPL and MATLAB.

4.2 Scenario one

First of all, MATLAB is used to create uniform random distributed subscribers in the coverage. The code of generating topology is as follow:

```
% 10 wifi 100 Mbps, 1 Macro LTE 100 Mbps
% R=600 meters R_wifi=100 meters
% 30 subscribers
clear all
close all

alfa = 0 : pi/20 : 2 * pi;
R=600;
R_w=100;

N_scr=30;% number of subscribers
r1=500;
r2=400;
r3=300;
r4=200;
r5=100;

% coordinator of Macro LTE
X=600; Y=600;
% coordinator of WiFi
x1=100;y1=600;x2=460;y2=170;
x3=375;y3=950;x4=270;y4=400;
x5=425;y5=625;x6=875;y6=700;
x7=725;y7=975;x8=650;y8=400;
x9=950;y9=925;x10=975;y10=325;

for n=1:N_scr
    m1(n)=R * sqrt(rand(1,1));
    theta1(n)=2 * pi * rand(1,1));
    P1(n)=X + m1(n) * cos(theta1(n));
    Q1(n)=Y + m1(n) * sin(theta1(n));
end
```

```
h1=plot(X,Y,'r*',X + R * cos(alfa), Y + R * sin(alfa),'k','LineWidth',2);
text(X,Y,'LTE')
text(X - 30, Y - 50,'QAM - 64')
hold on;
h2=plot(x1,y1,'g*', x1 + R_w * cos(alfa), y1 + R_w * sin(alfa));
text(x1,y1,'1');
plot(x2,y2,'g*', x2 + R_w * cos(alfa), y2 + R_w * sin(alfa));
text(x2,y2,'2');
plot(x3,y3,'g*', x3 + R_w * cos(alfa), y3 + R_w * sin(alfa));
text(x3,y3,'3');
plot(x4,y4,'g*', x4 + R_w * cos(alfa), y4 + R_w * sin(alfa));
text(x4,y4,'4');
plot(x5,y5,'g*', x5 + R_w * cos(alfa), y5 + R_w * sin(alfa));
text(x5,y5,'5');
plot(x6,y6,'g*', x6 + R_w * cos(alfa), y6 + R_w * sin(alfa));
text(x6,y6,'6');
plot(x7,y7,'g*', x7 + R_w * cos(alfa), y7 + R_w * sin(alfa));
text(x7,y7,'7');
plot(x8,y8,'g*', x8 + R_w * cos(alfa), y8 + R_w * sin(alfa));
text(x8,y8,'8');
plot(x9,y9,'g*', x9 + R_w * cos(alfa), y9 + R_w * sin(alfa));
text(x9,y9,'9');
plot(x10,y10,'g*', x10 + R_w * cos(alfa), y10 + R_w * sin(alfa));
text(x10,y10,'10');

plot(X,Y,X + r1 * cos(alfa), Y + r1 * sin(alfa),'k');
text(X - 595, Y, 'QPSK');
plot(X,Y,X + r2 * cos(alfa), Y + r2 * sin(alfa),'k');
text(X - 495, Y, 'QPSK');
plot(X,Y,X + r3 * cos(alfa), Y + r3 * sin(alfa),'k');
text(X - 395, Y, 'QAM16');
plot(X,Y,X + r4 * cos(alfa), Y + r4 * sin(alfa),'k');
text(X - 295, Y, 'QAM16');
plot(X,Y,X + r5 * cos(alfa), Y + r5 * sin(alfa),'k');
text(X - 195, Y, 'QAM64');
hold on;
h3=plot(P1,Q1,'m.');
```

legend([h1(1),h2(1),h3(1)], 'Macro LTE', 'Wifi', 'Subscriber');
axis equal;
hold off;
%axix off;

After running the MATLAB, we could get the first scenario's topology as in Figure 4.4.

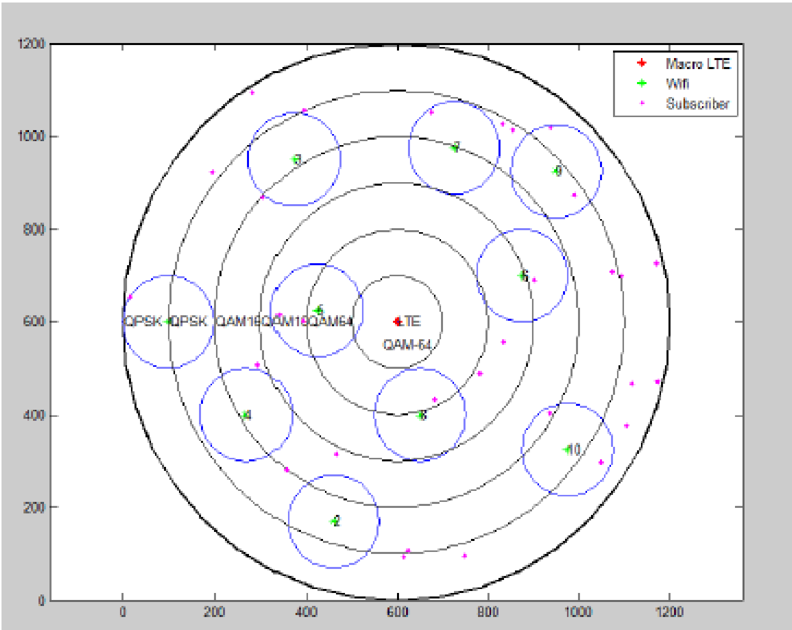

Figure 4.5: Topology of the scenario 1

As shown above, there are 30 subscribers in total. To make the topology more reasonable, these subscribers are uniformly random distributed. Randomly, 10 of them are found inside of Wi-Fi coverage in the first scenario. Also, there are 6 sub-circles are created to divide the whole coverage into 6 different radiuses' sub-circles. The radiuses of 6 circles are 600,500,400,300,200 and 100 meters, respectively. The spectral efficiencies of each sub-circle (from outside to inside) are 1,1.5,2,3,4 and 4.5 bit/s/Hz since the MCSs are QPSK, QAM-16 and QAM-64.Obviously, 13 subscribers are in the first QPSK sub-circle (1bit/s/Hz), 7 subscribers are in the second QPSK sub-circle (1.5 bit/s/Hz), 5 subscribers are in the first QAM-16 sub-circle (2 bit/s/Hz), 4 subscribers are in the second QAM-16 sub-circle (3 bit/s/Hz), 1 subscriber is in the first QAM-64 sub-circle and none subscriber is in the second QAM-64 (4.5 bit/s/Hz). As shown in Table 4.1.

Spectral efficiency	1	1.5	2	3	4	4.5
The number of subscriber	13	7	5	4	1	0

Table 4.1: Spectral efficiency of subscribers in scenario 1

In the topology of scenario 1, there are 1 macro-LTE base station in the center marked with red point. And, 10 Wi-Fi base stations are in the specific positions marked with green points. In MATLAB, we set the fixed positions for 10 Wi-Fi base stations in the coordinate system. Besides, 30 subscribers are marked with pink points. The bandwidths of both macro-LTE and Wi-Fi are set to be 20 MHz.

The relation between nodes and paths is shown in Table 4.2.

Wi-Fi	Spectral efficiency(bit/s/Hz)	Node	Path
1	1	1	2
5	3	26	6
5	3	27	6
6	2	25	7
7	1.5	17	8
8	4	30	9
9	1	9	10
9	1.5	15	10
10	1	5	11
10	2	21	11

Table 4.2: Relation between nodes and paths in scenario 1.

For instance, there are two subscribers inside of the tenth Wi-Fi coverage.One of them is also in the first QPSK sub-circle, which means the spectral efficiency of this node is 1 bit/s/Hz. Another node is in the first QAM-16 sub-circle, which means the spectral efficiency of this node is 2 bit/s/Hz. From node 1 to node 30, we set the spectral efficiency of each one sequentially. In other words, we set node 1 to node 13 all have spectral efficiency of 1 bit/s/Hz, and node 21 to node 25 all have spectral efficiency of 2 bit/s/Hz. Then, we let two points which are inside of Wi-Fi 10 coverage are node 5 and node 21. Since the path 1 always means the path connecting to macro-LTE base station. The path 11 is the path connecting to Wi-Fi 10 base station. Besides, node 5 and node 21 share the same path to Wi-Fi 10, so both of them use path 11.

Now, we could create data file in AMPL, and set the Path[d,p] as in the Figure4.5.

```
set Path[1,1] := 1; set Path[1,2] := 2;
set Path[2,1] := 1;
set Path[3,1] := 1;
set Path[4,1] := 1;
set Path[5,1] := 1; set Path[5,2] := 11;
set Path[6,1] := 1;
set Path[7,1] := 1;
set Path[8,1] := 1;
set Path[9,1] := 1; set Path[9,2] := 10;
set Path[10,1] := 1;
set Path[11,1] := 1;
set Path[12,1] := 1;
set Path[13,1] := 1;
set Path[14,1] := 1;
set Path[15,1] := 1; set Path[15,2] := 10;
set Path[16,1] := 1;
set Path[17,1] := 1; set Path[17,2] := 8;
set Path[18,1] := 1;
set Path[19,1] := 1;
set Path[20,1] := 1;
set Path[21,1] := 1; set Path[21,2] := 11;
set Path[22,1] := 1;
set Path[23,1] := 1;
set Path[24,1] := 1;
set Path[25,1] := 1; set Path[25,2] := 7;
set Path[26,1] := 1; set Path[26,2] := 6;
set Path[27,1] := 1; set Path[27,2] := 6;
set Path[28,1] := 1;
set Path[29,1] := 1;
set Path[30,1] := 1; set Path[30,2] := 9;
```

Figure 4.6: Part of data file in AMPL in scenario 1

Next, after running the AMPL script file, we could obtain the optimal solution h[d] as in the Figure 4.6. To verify the solution, we also obtain the x[d,p] at the same time as shown in Figure 4.7.

```
The optimal solution is below:
h [*] :=
 1     2e+07       9       1.2e+07   17     3e+07       25     4e+07
 2  1290320       10  1290320        18  1290320        26     3e+07
 3  1290320       11  1290320        19  1290320        27     3e+07
 4  1290320       12  1290320        20  1290320        28  1290320
 5 13333300       13  1290320        21 13333300        29  1290320
 6  1290320       14  1290320        22  1290320        30     8e+07
 7  1290320       15       1.2e+07   23  1290320
 8  1290320       16  1290320        24  1290320
```

Figure 4.7: Optimal solution h[d] of the scenario 1

```
x [*,*]
:              1                  2            :=
1           0               2e+07
2     1290320                    .
3     1290320                    .
4     1290320                    .
5           0            13333300
6     1290320                    .
7     1290320                    .
8     1290320                    .
9           0                    1.2e+07
10    1290320                    .
11    1290320                    .
12    1290320                    .
13    1290320                    .
14     860215                    .
15          0               8e+06
16     860215                    .
17          0               2e+07
18     860215                    .
19     860215                    .
20     860215                    .
21          0             6666670
22     645161                    .
23     645161                    .
24     645161                    .
25          0               2e+07
26          0               1e+07
27          0               1e+07
28     430108                    .
29     430108                    .
30         -2.32831e-09     2e+07
;
```

Figure 4.8: x[d,p] of the scenario 1

In order to intuitively illustrate the optimal solution, we convert the optimal solution into bar chart using MATLAB. As shown in Figure 4.8 below.

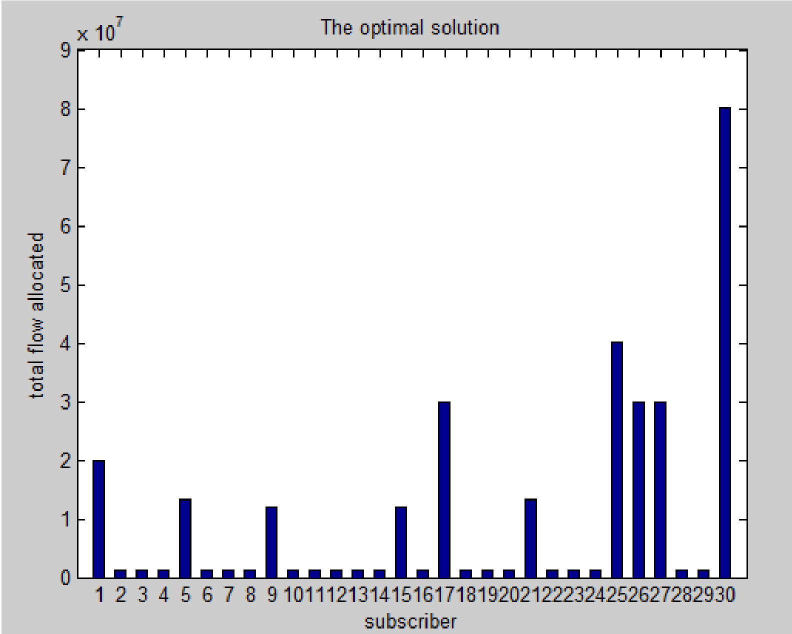

Figure 4.9: Bar chart of optimal solution of scenario 1

Then, more clear bandwidth allocation could be found in the figure. As can be seen from the bar chart, the bandwidth assigned to subscriber 30 reaches the highest point at $8 * 10^7$, since the spectral efficiency of subscriber 30 is 4 bit/s/Hz, and subscriber 30 occupies a single path. Meanwhile, the spectral efficiency of both subscriber 15 and subscriber 17 are 1.5 bit/s/Hz. However, subscriber 17 occupy path 8 alone while subscriber 15 share path 10 with subscriber 9. Therefore, the bandwidth assigned to subscriber 17 are almost twice larger than that of subscriber 15.

At the same time, we could find that the bandwidth assigned to subscriber 5 and 9 is different. However, the spectral efficiency of both of them is the same. As we can see, subscriber 9 share the same path with subscriber 15. And, the spectral efficiency of subscriber 15 is 1.5 bit/s/Hz. While, the subscriber 5 share the path 11 with subscriber 21, and the spectral efficiency of subscriber 21 is 2 bit/s/Hz. Hence, the bandwidth assigned to subscriber 5 is larger than that of subscriber 9. After that, we could find the bandwidth assigned to subscriber 5 and 21remains the same from the Figure 4.9.In other words, two subscriber would have the same bandwidth if they share the same path connecting to Wi-Fi base station and macro-LTE base station.

In comparison, most of the subscribers are not inside of Wi-Fi coverage, so they all have one path connecting to macro-LTE base station, which means they share the bandwidth of macro-LTE base station. The bandwidth assigned to them reaches the bottom at 1290320 Hz.

4.3 Scenario two

Similar with the first scenario, we also create a topology in MATLAB as in Figure 4.9.

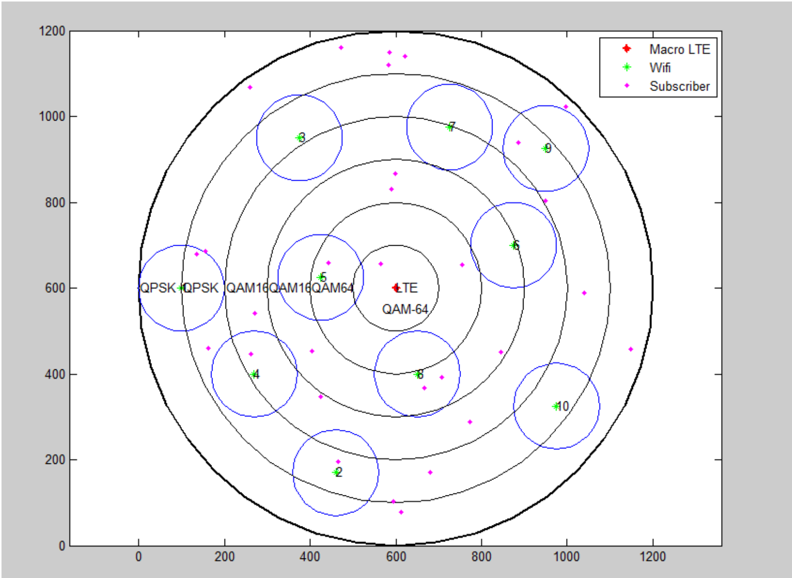

Figure 4.10: Topology of the scenario 2

As shown in the topology, 30 subscribers also uniformly random distributed. The spectral efficiencies of 6 sub-circles remain the same with the scenario 1, which are still 1,1.5,2,3,4 and 4.5 bit/s/Hz. Nevertheless, there are 8 subscribers in the first QPSK sub-circle (1 bit/s/Hz), 9 subscribers in the second QPSK sub-circle (1.5 bit/s/Hz). Besides, 4 subscribers are in the first QAM-16 sub-circle (2 bit/s/Hz), 6 subscribers are in the second QAM-16 sub-circle (3 bit/s/Hz), 2 subscribers are in the first QAM-64 sub-circle (4 bit/s/Hz), and only 1 subscriber is in the second QAM-64 sub-circle (4.5 bit/s/Hz). As shown in Table 4.3.

Spectral efficiency	1	1.5	2	3	4	4.5
The number of subscriber	8	9	4	6	2	1

Table 4.3: Spectral efficiency of subscribers in scenario 2

The positions of 10 Wi-Fi base station remain the same, and 7 subscribers are inside of Wi-Fi coverage. The bandwidth of both macro-LTE base station and Wi-Fi base station is still 20 MHz.

And, the relation between nodes and paths is shown in Table 4.4.

Wi-Fi	Spectral efficiency(bit/s/Hz)	Node	Path
1	1.5	9	2
2	1.5	15	3
4	2	21	5
5	4	27	6
8	3	22	9
8	3	25	9
9	1.5	17	10

Table 4.4: Relation between nodes and paths in scenario 2.

According to Table 4.4, there are 7 subscribers inside of Wi-Fi coverage in this scenario. Besides, node 22 and node 25 share path 9. Node 22 and node 25 also have the same spectral efficiency. Other nodes which are inside of Wi-Fi coverage all have two paths connecting to macro-LTE base station and Wi-Fi base station separately. Then, we could create the data file in AMPL as in Figure 4.10.

```
set Path[1,1] := 1;
set Path[2,1] := 1;
set Path[3,1] := 1;
set Path[4,1] := 1;
set Path[5,1] := 1;
set Path[6,1] := 1;
set Path[7,1] := 1;
set Path[8,1] := 1;
set Path[9,1] := 1; set Path[9,2] := 2;
set Path[10,1] := 1;
set Path[11,1] := 1;
set Path[12,1] := 1;
set Path[13,1] := 1;
set Path[14,1] := 1;
set Path[15,1] := 1; set Path[15,2] := 3;
set Path[16,1] := 1;
set Path[17,1] := 1; set Path[17,2] := 10;
set Path[18,1] := 1;
set Path[19,1] := 1;
set Path[20,1] := 1;
set Path[21,1] := 1; set Path[21,2] := 5;
set Path[22,1] := 1; set Path[22,2] := 9;
set Path[23,1] := 1;
set Path[24,1] := 1;
set Path[25,1] := 1; set Path[25,2] := 9;
set Path[26,1] := 1;
set Path[27,1] := 1; set Path[27,2] := 6;
set Path[28,1] := 1;
set Path[29,1] := 1;
set Path[30,1] := 1;
```

Figure 4.11: Part of data file in AMPL in scenario 2

Then, we could run the script file in AMPL, and obtain the optimal solution. The result shows as in the Figure 4.11.

```
The optimal solution is below:
h [*] :=
 1 1313870    6 1313870   11 1313870   16 1313870   21    4e+07   26 1313870
 2 1313870    7 1313870   12 1313870   17    3e+07   22    3e+07   27    8e+07
 3 1313870    8 1313870   13 1313870   18 1313870   23 1313870   28 1313870
 4 1313870    9    3e+07  14 1313870   19 1313870   24 1313870   29 1313870
 5 1313870   10 1313870   15    3e+07  20 1313870   25    3e+07   30 1313870
;
```

Figure 4.12: Optimal solution h[d] of the scenario 2

We also obtain bandwidth allocation x[d,p] to verify the solution as in the Figure 4.12.

```
x [*,*]
:           1             2      :=
1     1313870             .
2     1313870             .
3     1313870             .
4     1313870             .
5     1313870             .
6     1313870             .
7     1313870             .
8     1313870             .
9          -2.48353e-09   2e+07
10     875912             .
11     875912             .
12     875912             .
13     875912             .
14     875912             .
15          0             2e+07
16     875912             .
17          0             2e+07
18     656934             .
19     656934             .
20     656934             .
21         -3.72529e-09   2e+07
22          0             1e+07
23     437956             .
24     437956             .
25          0             1e+07
26     437956             .
27          4.65661e-10   2e+07
28     328467             .
29     328467             .
30     291971             .
;
```

Figure 4.13: x[d,p] of the scenario 2

To make the optimal solution more intuitive, we also convert the result into the bar chart in MATLAB. The optimal solution is illustrated as in the Figure 4.13.

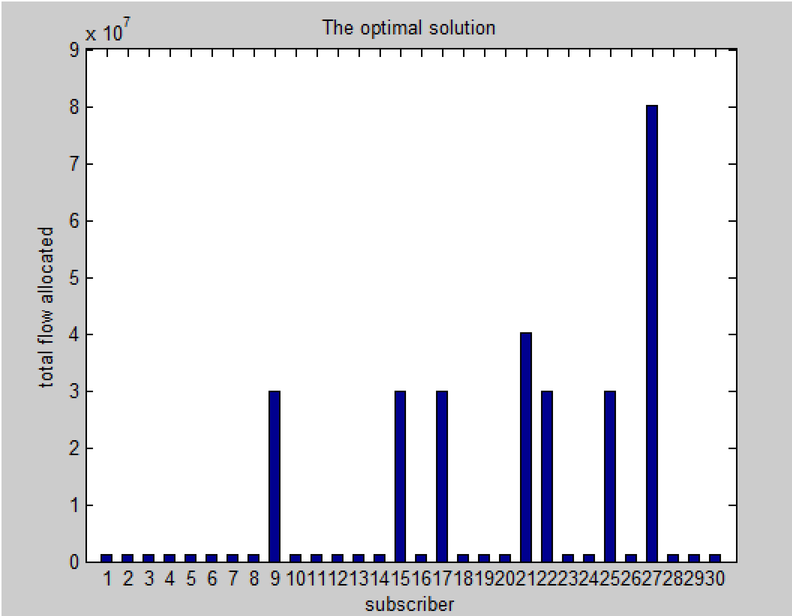

Figure 4.14: Bar chart of optimal solution of scenario 2

We can see from the bar chart that the bandwidth assigned to subscriber 27 reaches the peak at $8 * 10^7$. Firstly, subscriber 9,15 and 17 have the same spectral efficiency, and they all have single path connecting to Wi-Fi base station. That is the reason why the bandwidth assigned to them is totally the same. Obviously, the bandwidth of subscriber 21 is the second highest in the Figure 4.13. It is because that the spectral efficiency of subscriber 21 is 2 bit/s/Hz, besides, subscriber 21 occupy path 5 alone. In the second scenario, only subscriber 22 and 25 are inside of the same Wi-Fi coverage, so they need to share the bandwidth of Wi-Fi 8. As shown in the Table 4.4, the spectral efficiency of both subscriber 22 and 25 is twice larger than that of subscriber 9,15 and 17, so the bandwidth of them turns out to be equal in the optimal solution.

Except the subscribers mentioned above, other subscribers are not inside of any Wi-Fi coverage, it means they all have only one path connecting to macro-LTE base station. They divide the bandwidth left equally, and the bandwidth assigned to each of them is 1313870Hz.

4.4 Scenario three

To make sure the validity of the algorithm, another scenario is created. Similar with the first two scenarios, we need to uniformly random distribute 30 subscribers within the coverage. As shown the topology in Figure 4.14.

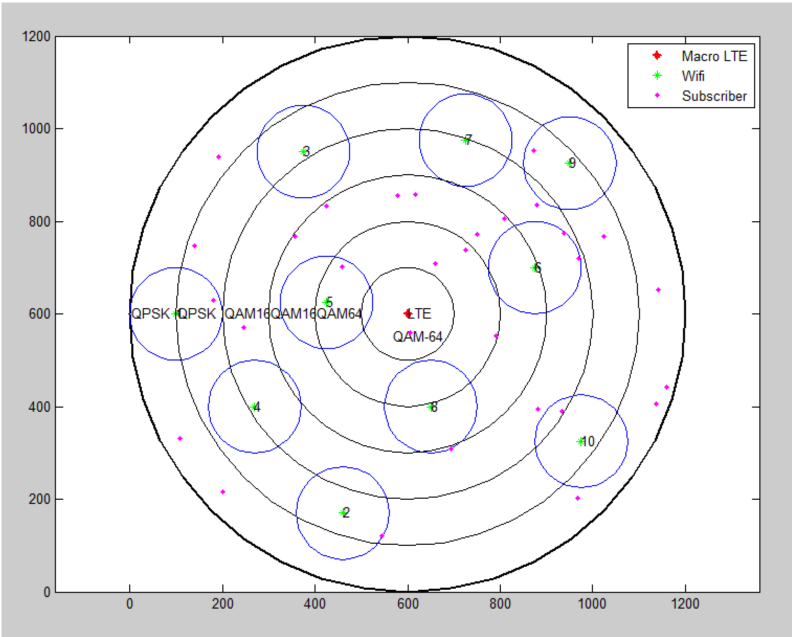

Figure 4.15: Topology of the scenario 3

As can be seen from the topology, the distribution of 30 subscribers are random. The spectral efficiencies of 6 sub-circles remain the same with the first two scenarios, which are still 1,1.5,2,3,4 and 4.5 bit/s/Hz. In this scenario, we have 7 subscribers in the first QPSK sub-circle (1 bit/s/Hz), and 5 subscribers in the second QPSK sub-circle (1.5 bit/s/Hz). Also, 7 subscribers are in the first QAM-16 sub-circle (2 bit/s/Hz), 6 subscribers are in the second QAM-16 sub-circle (3 bit/s/Hz), 4 subscribers are in the first QAM-64 sub-circle (4 bit/s/Hz), and only 1 subscriber is in the second QAM-64 sub-circle (4.5 bit/s/Hz). As shown in the Table 4.5.

Spectral efficiency	1	1.5	2	3	4	4.5
The number of subscriber	7	5	7	6	4	1

Table 4.5: Spectral efficiency of subscribers in scenario 3

The positions of 10 Wi-Fi base stations are not changed, and 8 subscribers are connected to Wi-Fi base stations. The bandwidth of macro-LTE and Wi-Fi base stations is still 20 MHz.

Furthermore, the relation between nodes and paths is shown in Table 4.6.
Based on Table 4.6, it has 8 subscribers connecting Wi-Fi base stations in scenario

Wi-Fi	Spectral efficiency(bit/s/Hz)	Node	Path
1	1.5	8	2
2	1.5	9	3
5	4	27	6
6	2	14	7
6	2	15	7
8	2	17	9
9	1.5	11	10
10	2	19	11

Table 4.6: Relation between nodes and paths in scenario 3.

3. For instance, node 14 and node 15 both connect to Wi-Fi 6, which means they share path 7. While node 17 occupy path 10 connecting to Wi-Fi 8. It should be noticed that node 14,15 and 17 have the same spectral efficiency. Hence, we could create data file as in Figure 4.15.

```
set Path[1,1] := 1;
set Path[2,1] := 1;
set Path[3,1] := 1;
set Path[4,1] := 1;
set Path[5,1] := 1;
set Path[6,1] := 1;
set Path[7,1] := 1;
set Path[8,1] := 1; set Path[8,2] := 2;
set Path[9,1] := 1; set Path[9,2] := 3;
set Path[10,1] := 1;
set Path[11,1] := 1; set Path[11,2] := 10;
set Path[12,1] := 1;
set Path[13,1] := 1;
set Path[14,1] := 1; set Path[14,2] := 7;
set Path[15,1] := 1; set Path[15,2] := 7;
set Path[16,1] := 1;
set Path[17,1] := 1; set Path[17,2] := 9;
set Path[18,1] := 1;
set Path[19,1] := 1; set Path[19,2] := 11;
set Path[20,1] := 1;
set Path[21,1] := 1;
set Path[22,1] := 1;
set Path[23,1] := 1;
set Path[24,1] := 1;
set Path[25,1] := 1;
set Path[26,1] := 1;
set Path[27,1] := 1; set Path[27,2] := 6;
set Path[28,1] := 1;
set Path[29,1] := 1;
set Path[30,1] := 1;
```

Figure 4.16: Part of data file in AMPL in scenario 3

Next, we could obtain the optimal solution after running the script file in AMPL. The optimal solution h[d] is shown as in Figure 4.16.

```
The optimal solution is below:
h [*] :=
 1 1561820    6 1561820   11    3e+07   16 1561820   21 1561820   26 1561820
 2 1561820    7 1561820   12 1561820   17    4e+07   22 1561820   27    8e+07
 3 1561820    8    3e+07  13 1561820   18 1561820   23 1561820   28 1561820
 4 1561820    9    3e+07  14    2e+07  19    4e+07   24 1561820   29 1561820
 5 1561820   10 1561820   15    2e+07  20 1561820   25 1561820   30 1561820
;
```

Figure 4.17: Optimal solution h[d] of the scenario 3

Likewise, bandwidth allocation x[d,p] is needed to verify the optimal solution as in the Figure 4.17.

```
x [*,*]
:            1              2      :=
1     1561820               .
2     1561820               .
3     1561820               .
4     1561820               .
5     1561820               .
6     1561820               .
7     1561820               .
8          -2.48353e-09    2e+07
9           0              2e+07
10    1041210               .
11          0              2e+07
12    1041210               .
13     780911               .
14          0              1e+07
15          0              1e+07
16     780911               .
17          0              2e+07
18     780911               .
19          0              2e+07
20     520607               .
21     520607               .
22     520607               .
23     520607               .
24     520607               .
25     520607               .
26     390456               .
27         -4.65661e-10    2e+07
28     390456               .
29     390456               .
30     347072               .
;
```

Figure 4.18: x[d,p] of the scenario 3

The bar chart of the optimal solution is more easy to understand, so we use MATLAB to convert the solution to a bar chart as shown in the Figure 4.18.

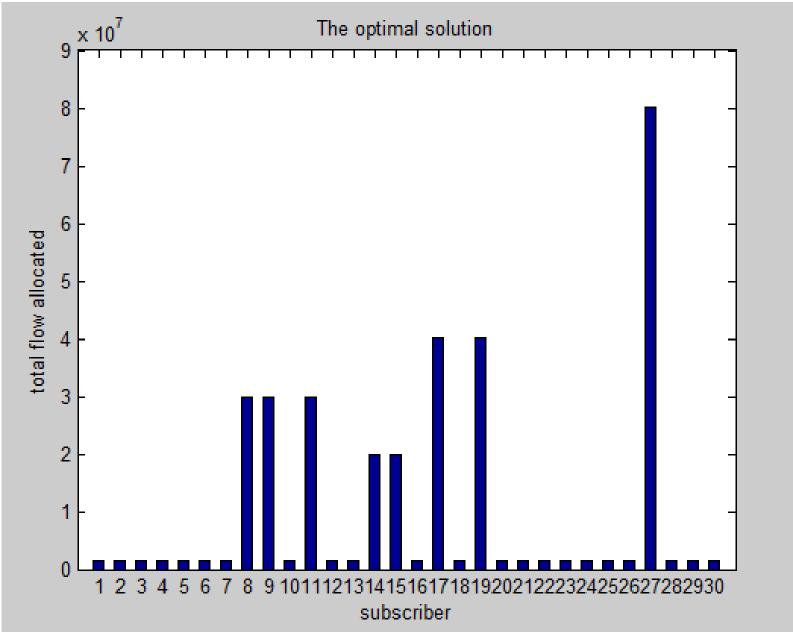

Figure 4.19: Bar chart of optimal solution of scenario 3

As mentioned above, the largest bandwidth allocation belongs to subscriber 27, since the spectral efficiency of it is 4 bit/s/Hz. Not only the bandwidth provided by macro-LTE base station, but also by Wi-Fi 5 base station. Hence, subscriber 27 is assigned with $8 * 10^7$Hz bandwidth.

After that, we should analyze other higher bandwidth allocations. In the bar chart, the bandwidth of subscriber 8, 9 and 11 remains the same because of the same spectral efficiency. Meanwhile, the spectral efficiency of subscriber 14,15,17 and 19 is 2 bit/s/Hz. Since, the subscriber 14 and 15 share the path 7 connecting to Wi-Fi 6, the bandwidth allocation is twice smaller than that of subscriber 17 and 19. It also could be found in the final optimal solution shown in Figure 4.17. It further proves the correctness of the algorithm.

Eventually, remaining subscribers all share the same path connecting to macro-LTE base station. Therefore, the bandwidth allocation demonstrated in Figure 4.19 is extremely low.

5. CONCLUSIONS

The ultimate objective of this thesis is to implement fair and priority-based rate allocation in heterogeneous cellular system using Max-Min fairness criterion. The concept of HetNet and cellular system is briefly introduced at the beginning of this thesis, and then current state and motivations of this topic is provided. Thereafter, we should realize there are still some problems existing currently.

As discussed in chapter 2, a simple illustration of network flow problem is briefly described and explained. Meanwhile, we introduce two notations of network flow problem and choose the better one to denote the flow variables. Then, we introduce the fairly important part, which is the criteria of fairness. MMF and PF are introduced and analyzed in details. In the last part of chapter 2, the previous study of the topic is described to make us realize some possible improvements.

In the most important theoretical part, chapter 3, the detailed description of rate allocation and formulation is fully presented. We demonstrate a particular heterogeneous system and resource allocation. In the next part, we illustrate the topology modeling and express the network flow formulation. Eventually, we propose an algorithm to apply it to spectral efficiency and priority.

In order to prove the correctness of the proposed algorithm, we create 3 different scenarios to simulate the realistic networks. During each implementation, we use MATLAB and AMPL to obtain the optimal solution. After analyzing the results, the method and algorithm is further tested.

According to the implementation simulation results presented in chapter 4, we successfully introduce two coefficients in the proposed algorithm. Consequently, it accomplish the goal of resource allocation in terms of spectral efficiency and prioritization in heterogeneous cellular networks.The rate assigned to each subscriber is fairly allocated based on MMF criterion. More importantly, prioritization could attract company's attention for applying into marketing. For instance, company could use the prioritization accomplished in this thesis to make various pricing schemes. Those various pricing schemes could give customers greater appeal, which finally brings higher profits.

Whereas, there are still some factors needed to taken into consideration in the future. The scenario is the most simplified ones in order to test the algorithm. It is not difficult to transfer it into the realistic one. It is because the routing could be implemented at macro-LTE base station, and invoke whenever something changes in the network. More importantly, it is necessary to realize that interference is existing all the time during the transmission.

To conclude, this chapter has reviewed what has been introduced in this thesis, described what needs to be further improved on this topic. The principle solving heterogeneous networks problem with MMF has filled the gap in this area. It will definitely attract attention to this subject, and accelerate the development in the future.

REFERENCES

[1] S.Landstrom,K.Johansson, ”Heterogeneous Networks(HetNets) – an approach to increasing cellular capacity and coverage”.
Available: http://www.ericsson.com/news/110211_hetnets_244188811_c, Accessed 03 November,2014.

[2] R.Periannan, ”Existing Technology of Cellular Networks-Mobile Phones”, Imperial College UK.

[3] T.Bhandare,”LTE and WiMAX Comparison”, Santa Clara University, December,2008.

[4] M.Rahman, F.A.M.Mir,”Fourth Generation(4G) Mobile Networks-Features, Technologies & Issues”, Cable & Wireless, Small House,UK.

[5] ”5G: A Technology Vision”,Huawei,2013.

[6] D.Moltchanov et al., ”On the Optimal Assisted Rate Allocation in N-tier Multi-RAT Heterogeneous Networks,” IEEE PIMRC 2014,USA.

[7] A.Khandekar, ”LTE-Advanced: Heterogeneous Networks,” European Wireless Conference, 2010.

[8] V.Chan et al., ”Workshop on Future Heterogeneous Networks,” 2011.

[9] M.Pioro, D.Mehdi, Routing, Flow, and Capacity Design in Communication and Computer Networks (The Morgan Kaufmann Seriesin Networking). 1 ed., 2004.

[10] Q.Wang, ”On the Optimal Assisted Rate Allocation in N-Tier Multi-RAT Heterogeneous Networks,” Auguest,2014.

[11] Fourer,R.,Brain W.Kernighan.,2002.AMPL:A Modeling Language for Mathematical Programming. Duxbury Press Addison-Wesley. 1120 s.

[12] LTE overview, Available: http://www.tutorialspoint.com/lte/lte_overview.htm, accessed 09 December 2014.

[13] Long Term Evolution Protocol, Available: http: www.freescale.com, accessed 30 December 2014.

[14] Mobile Phones: Jargon Explained, Avaiable:http://stakeholders.ofcom.org.uk/sitefinder/glossary/jargon/, accessed 15 December 2014.

[15] J.Zaloker,S.Weeres, 802.11a/b/g/n Wi-Fi, Arrow Electronics.

[16] IEEE 802.11a-1999

[17] IEEE SA-802.11b-1999

[18] IEEE SA-802.11g-2003

[19] IEEE SA-802.11n-2009

[20] IEEE SA-802.11ac-2013